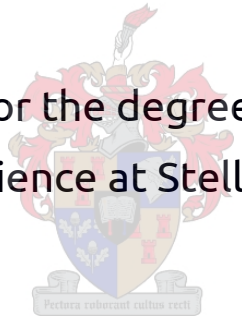
Identification and genomic distribution of select combinatorial post-translational modification patterns of *Trypanosoma brucei* histones.

by

Johannes Petrus Maree



Dissertation presented for the degree of Doctor of Philosophy

in the Faculty of Science at Stellenbosch University

Promotor: Prof. Hugh-G. Patterton

Co-Promotor: Prof. Ann Louw

Department of Biochemistry,

March 2020

DECLARATION

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated) , that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

March 2020

Abstract

Trypanosoma brucei is a dixenous parasitic protozoan, and the causative agent of African Sleeping sickness, or Human African Trypanosomiasis. Trypanosomes diverged very early on from the main eukaryotic lineage and display some very unusual genomic features. Pol II transcribed genes are arranged in long, non-overlapping polycistronic transcription units (PTUs) containing up to a hundred functionally unrelated genes. These PTUs are separated by strand switching regions (SSRs) which can either be divergent (dSSR) or convergent (cSSR), depending on the direction of transcription. The entire housekeeping portion on the genome is constitutively transcribed with pol II initiating from poorly defined promoters. Putative transcription start sites are epigenetically demarcated by histone post-translational modifications (PTMs) and variant deposition, with transcription initiating bi-directionally from divergent SSRs as well as internal stop/start sites (ISS) at PTU head-to-tail regions.

Over the past two decades it has become increasingly clear that in *Trypanosoma* the epigenome has a vital role in gene regulation and transcriptional control. *T. brucei* lacks pol II promoter and other regulatory sequences and employs epigenetic marks to demarcate and control genomic function.

More than a decade has passed since the last survey of *T. brucei* histone PTMs, and improvements in mass spectrometry (MS) technology and software now allows us to map PTMs with great accuracy. Application of

these technologies revealed numerous histone PTMs in *T. brucei*, with some forming intricate combinatorial PTM patterns. The C-terminal tail of histone H2A was found to be hyperacetylated. Half of all H2A peptides observed were hyperacetylated, with up to 5 acetylated lysines occurring simultaneously. On histone H3 the lysine 10 residue was found to be tri-methylated, present as a binary PTM with hydroxylated H3P40. H3HyPro40 and H2AHyPro26 were the only hydroxylated residues detected, and further analysis supports the validity of this PTM in *T. brucei*.

Investigation of the genome-wide distribution of hyperacetylated H2A PTMs via MNase-ChIP-seq revealed that hyperacetylation of the C-terminal tail in specific forms served distinct functions and displayed different genomic localization. It was further determined that one of these states of hyperacetylation, H2AK125ac, was enriched at dSSRs and ISSs. The H2AK125ac pattern was seen to co-localize with H2A.V and RPB9, indicators of putative pol II transcription start sites. Immunofluorescence microscopy (IF) microscopy showed that different H2A hyperacetylation patterns localized to distinct nuclear foci, further emphasizing distinct functionality.

The data and findings of this study provides further insight into the epigenetic mechanisms employed by *T. brucei* and hopefully leads to development of targeted epigenetic therapies.

Opsomming

Trypanosoma brucei is 'n twee-fasige parasitiese protosoa, en die oorsaak van Afrika-slaapsiekte, of Menslike Afrika Trypanosomiasis. Trypanosome het baie vroeg van die hoof eukariotiese stam vertak en vertoon 'n paar baie ongewone genomiese kenmerke. Pol II-transkribeerde gene word rangskik in lang, nie-oorvleuelende polisistroniese transkripsie-eenhede (PTU's) wat tot honderd funksioneel onverwante gene bevat. Hierdie PTU's word geskei deur ketting-wisselareas (SSRs) wat óf divergent (dSSR) of konvergent (cSSR) kan wees, afhangende van die rigting van transkripsie. Die hele huishoudelike geen kompliment van die genoom word voltyds getranskribeer deur pol II wat by swak gedefinieerde promotors begin. Transkripsie begin areas word epigeneties afgebaken deur histoon PTMs en variant neervoeging, en transkripsie begin in twee rigting by divergente SSRs sowel as by interne stop / begin plekke (ISS) by PTU kop-na-stert gebiede.

Oor die afgelope twee dekades het dit al hoe duidelikker geword dat die trypanosomale epigenoom 'n deurslaggewende rol speel in transkripsionele beheer en geenregulering. *T. brucei* ontbreek pol II promotor en ander regulerende DNA elemente en gebruik epigenetiese merke om die genomiese funksie af te baken en te beheer. Meer as 'n dekade is verby sedert die laaste opname van *T. brucei* histoon PTMs. Verbetering in MS-tegnologie en sagteware laat ons nou toe om PTM's met groot akkuraatheid te karteer. Toepassing van hierdie tegnologieë het talle histoon PTM's in *T. brucei* getoon, met die vorming van ingewikkelde kombinatoriese PTM patrone. Die C-terminale stert van

H2A is bevind om gehiper-asetileer te wees. Die helfte van alle H2A-peptiede waargeneem is ge-asetileerd, met tot 5 ge-asetileerde lisiene wat gelyktydig voorkom. Op histoon H3 was die lisien 10 aminosuur getri-metileer, teenwoordig as 'n binêre PTM, saam met gehidroksileerde H3P40. H3HyPro40 en H2AHyPro26 is die enigste gehidroksileerde residu's wat waargeneem was, en verdere analise ondersteun die geldigheid van hierdie PTM in *T. brucei*. Genoom-wye verspreiding van gehiper-asetileerde H2A PTMs via MNase-ChIP-seq het getoon dat die hiper-asetilering van die C-terminale stert, in spesifieke vorme, verskillende lokalisering en funksies vertoon. Daar is verder bepaal dat een van hierdie vorme, H2AK125ac, by dSSRs en ISS's verryk was. Die H2AK125ac-patroon was gesien om saam met H2A.V en RPB9 voor te kom, aanwysers van pol II transkripsie begin plekke. IF mikroskopie het getoon dat verskillende H2A-hiper-hiperasetilasiepatrone gelokaliseer word na afsonderlike kern fokus plekke, wat unieke funksionaliteit beklemtoon.

Die data en bevindings van hierdie studie lewer verdere insig tot die epigenetiese meganismes wat deur *T. brucei* gebruik word, en lei hopelik tot die ontwikkeling van geteikende epigenetiese terapieë.

Acknowledgments

To the scrupulous and accommodating experts, the thoughtful and supportive colleagues, the gurus, and leaders from various scientific fields across the world that hosted, assisted, supported, and advanced my understanding and appreciation of the intricate subtlety of natural sciences throughout my life.

Professor Hugh G. Patterton - for introducing me to the fascinating field of genomics, epigenetics, and bioinformatics.

Professor Ann Louw who welcomed me into the Africander-Louw-Verhoog laboratory. A special thanks to Drs. N. Verhoog and L. Wilkinson who made me feel at home.

Professor Gloria Rudenko, for accommodating me in her laboratory; her guidance and discussions, as well as the Rudenko lab members. Special thanks to Mr. James Budzack for his assistance with IF microscopy.

Professor Ole Jensen for his insight into histone PTM analysis, Dr. Andrey Tvardovskviy for his hands-on assistance and guidance with MS analysis.

Dr. Enock Matovu for *in vivo* culturing of BF *T. brucei.*

The H3Africa initiative for funding this study; receiving financial support from the NIH (USA) and Welcome Trust (UK).

To my almighty Creator, my amazing family, my astonishing mother, and my awesome friends for believing, supporting, guidance and encouragement.

Table of Contents

List of Figures

List of Tables

List of abbreviations

ACN	Acetonitrile
BES	Bloodstream Expression Site
BF	Bloodstream Form
BSA	Bovine Serum Albumin
cPTM	Combinatorial Post Translational Modification
cSSR	Convergent Strand Switching Region
dSSR	Divergent Strand Switching Region
ESAG	Expression Site Associated Gene
ETD	Electron-Transfer Dissociation
FCS	Fetal Calf Serum
HT	Head-to-tail
IF	Immunofluorescence microscopy
ISS	Internal stop/start region
LC	Liquid Chromatography
MBC	Megabase Chromosome
MS	Mass Spectrometry
N-ChIP	Native Chromatin Immunoprecipitation
NDRs	Nucleosome Depleted region
PBS	Phosphate Buffed Saline
PF	Procyclic Form
P4H	Prolyl-4 -hydroxylase
PTM	Post Translational Modification
PTU	Polycistronic Transcription Unit
RCF	Relative Centrifugal Force
RP-HPLC	Reverse-phase high-performance liquid chromatography
SAS	Splice Acceptor Site
SSR	Strand Switching Regions
TIC	Total Ion Chromatogram
TSS	Transcription Start Site
TTS	Transcription Termination Site
VSG	Variable Surface Glycoprotein
WCX-HILIC	Weak cation exchange hydrophobic interaction liquid chromatography

1) Introduction

It is generally accepted that the eukaryotic progenitor arose some 1.8 billion years ago (Knoll et al. 2006). The sudden expansion in cellular complexity necessitated a vast increase of genetic information encoded in DNA which, in turn, had to fit into the nucleus - a specialized cellular compartment which is, on average, 6 µm in diameter (Huber and Gerace 2007). This remarkable degree of compaction is accomplished by wrapping ~146 bp of DNA in two negative supercoils around a histone octamer to condense the polyanionic DNA and form a nucleosome, the basic structural unit of chromatin and an ancient evolutionary adaptation shared between Archaea and Eukarya (Luger et al. 1997; Ammar et al. 2012). Nucleosome formation is generally a transcriptionally repressive process as access to key DNA binding sites by transcriptional machinery is impeded. Access to genomic loci and DNA binding sites, chromatin compaction, and gene regulation is facilitated through histone variant deposition and chromatin remodelers. Chromatin remodelers are enzymes capable of altering chromatin states by condensing DNA into a repressive heterochromatic state, or inducing a relaxed euchromatic state amenable to transcription. Over the past few decades, numerous studies contributed to a vast increase in knowledge regarding complex epigenetic mechanisms that influence a wide range of processes; from regulation of genomic function in model eukaryotes, to influence on cognition and emotional states in humans (Talbert and Henikoff 2009; Akers et al. 2018).

The protist superclade Excavata, which encompasses the order Trypanosomatida, separated remarkably early from the main eukaryotic tree (Cavalier-Smith 2010). *Trypanosoma brucei* is an extracellular dixenous parasite of the mammalian bloodstream and the causative agent of Human African trypanosomiasis and Nagana in livestock. Estimates show that in sub-Saharan Africa approximately 60 million people are at risk of contracting this debilitating disease, resulting in roughly 1.3 million disability-adjusted life years lost (Fèvre et al. 2008; World Health Organization 2013). The mortality rate of Human African trypanosomiasis is close to 100% if left untreated (WHO 2015). The treatment course administered depends on the type and stage of infection, with chemotherapeutic agents like pentamidine and eflornithine often having unwanted and toxic side-effects on patients.

T. brucei has two main life cycle stages. One occurs in the insect stage, where trypanosomes transform to procyclic form (PF) trypomastigotes, multiplying by asexual cell division in the insect midgut (Fig. 1.1). *T. brucei* then migrates to the salivary gland of the insect vector (a fly from the *Glossina spp.*) and transforms to infective metacyclic epimastigotes. Upon a blood meal by the vector, trypanosomes are transferred via the insect salivary duct to a mammalian host where it transforms to the bloodstream form (BF) and multiplies by binary fission. During this transformation, the outer cell surface is covered in a dense layer of a single protein isoform known as the Variable Surface Glycoprotein (VSG). This dense coat of up to 1×10^7 VSGs per cell effectively shields invariable cell surface antigens from recognition by the host immune system.

T. brucei encodes a vast repertoire of ~2000 VSG genes that can recombine in mosaic patterns (Taylor and Rudenko 2006). Periodic switching of the expressed VSG by recombination reactions exhausts the immune response long before the infection can be cleared by the host immune system, leading to host mortality.

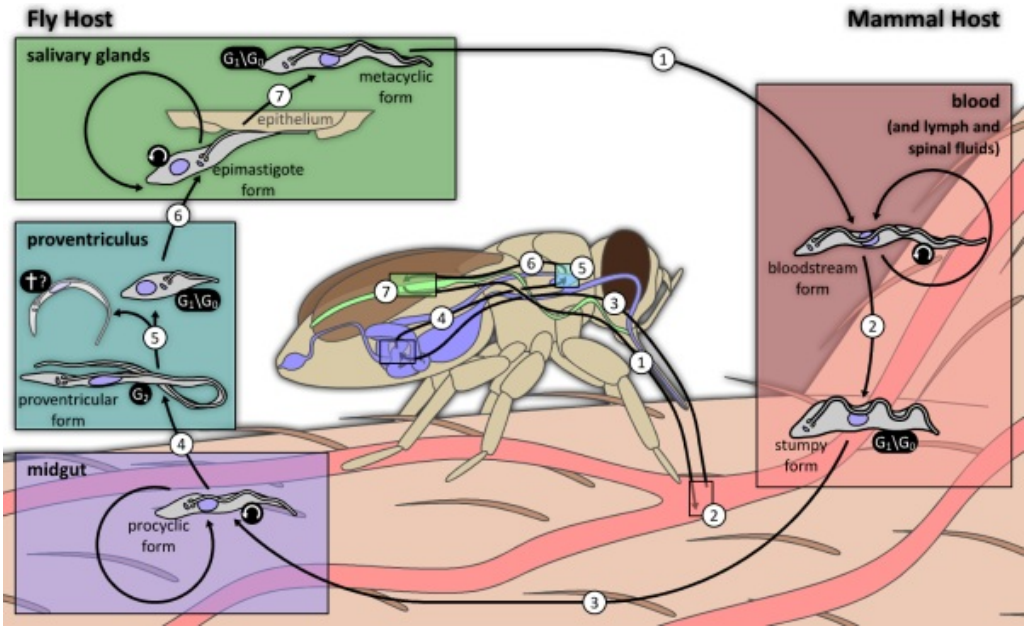


Figure 1.1. Life cycle of *Trypanosoma brucei*: (1) Infective metacyclic epimastigotes are introduced to the mammalian host during a blood meal from an infected Tsetse fly, transforming into BF trypanosomes. (2) BF cells transform into stumpy, non-dividing cells (3) capable of vector uptake and fly midgut infection (4) and transformation to PF trypanosomes. (5 - 7) Migration and transformation of PF cells to infective metacyclic epimastigotes in the salivary gland. Reproduced with permission (Wheeler, n.d.).

This early divergent parasite displays some unusual and unique genomic, transcriptomic, and epigenetic characteristics. Polymerase (pol) I, in addition to transcribing 5S rRNA genes, also transcribes the VSG protein at very high rates (Günzl et al. 2003). A VSG gene is transcribed from a

single, active polycistronic transcription unit (PTU), known as the bloodstream expression site (BES). There are 15 BES in *T. brucei* of which only 1 is transcriptionally active at a time (Navarro and Gull 2001). It is crucial for host immune evasion and parasite survival that the remaining 14 BESs are transcriptionally silenced, a regulatory process appearing to be predominantly epigenetic.

The housekeeping portion of the genome is arranged as long, non-overlapping gene arrays containing up to a hundred functionally unrelated genes (Daniels, Gull, and Wickstead 2010). These PTUs are constitutively transcribed by pol II, and pre-mRNAs are co-transcriptionally polyadenylated and trans-spliced. During trans-splicing, a 37 nt long RNA molecule known as the spliced leader (SL) RNA is added to the 5' side of every gene transcript at the splice acceptor site (SAS). This allows the translational machinery to translate proteins from transcripts originating from different polymerases.

The haploid genome size varies between ~26.5 to 35 Mb depending on the strain (Berriman et al. 2005; Daniels, Gull, and Wickstead 2010), caused by size variability of sister chromatids, non-Mendelian mini- and intermediate chromosome inheritance, and the intriguing feature of telomere elongation (Alsford et al. 2001). The nuclear genome consists of 11 mega-base chromosomes (MBCs), ~5 intermediate chromosomes (200 - 900 kb), and 50 - 200 mini-chromosomes (50 - 150 kb) (Daniels, Gull, and Wickstead 2010).

In addition to the above-mentioned genomic peculiarities, *T. brucei* lacks any detectable pol II activator or other regulatory DNA sequences, and gene regulation mostly occurs post-transcriptionally (Droll et al. 2013; Clayton 2019). Pol II PTUs are separated by strand switching regions (SSRs) which can either be divergent or convergent, depending on the direction of transcription (Fig. 1.2), and can also be arranged in a head-to-tail (HT) manner.

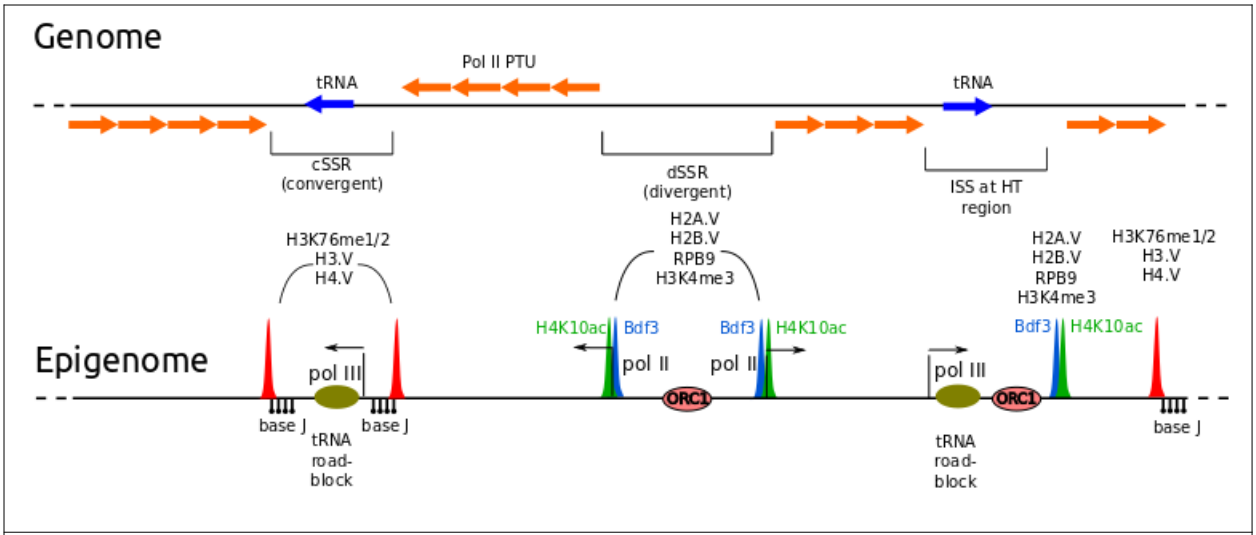


Figure 1.2: Genomic organization and epigenetic signals demarcating pol II PTUs, regulate transcription initiation and termination, and genome function. Individual genes are indicated by orange arrows.

With the exception of the SL RNA, which contain the only pol II promoter sequence, pol II PTUs are devoid of promoter sequences. Instead, pol II transcription initiates bi-directionally from divergent SSRs (dSSRs), and terminates at convergent SSRs (cSSRs). Transcription can also initiate at HT regions, known as internal stop/start sites (ISSs). These sites are frequently populated by pol III transcribed genes (like tRNAs) and active pol III transcription at these ISS sites presents a steric hindrance to a

transcribing pol II which would need to terminate and re-initiate downstream of such a tRNA "roadblock".

A myriad of epigenetic marks, readers and writers have been discovered in *T. brucei*, some of which have been mapped to specific, and often crucial functions (Fig. 1.2). Histone variants H2A.V and H2B.V, histone PTMs like acetylated H4K10 and tri-methylated H3K4, the chromatin remodeler BDF3, and the pol II sub-unit RPB9 were found to co-localize at putative pol II transcription start sites (TSS) (Siegel et al. 2009; Wright, Siegel, and Cross 2010). At regions of transcription termination, histone variants H3.V and H4.V, and mono- or di-methylated H3K76 coincide at cSSRs.

Work by McAndrew and colleagues demonstrated that in *T. brucei* an open chromatin structure was sufficient to precipitate and initiate pol II transcription (McAndrew et al. 1998). It is, therefore, possible that the presence of histone variants and PTMs, and chromatin remodelers at putative TSSs contribute to a more euchromatic chromatin structure amenable to pol II transcription. Indeed, nucleosomal organization at pol-II PTU transcription initiation regions display a nucleosomal architecture similar to that seen in higher eukaryotes. A nucleosome depleted region (NDR) was observed at the first SAS of the first gene of a PTU, bordered by well-positioned -1 and +1 nucleosomes (Maree et al. 2017). In contrast to the first SASs, a striking enrichment of nucleosomal dyads were observed to be centered on all internal SASs.

Histones are some of the most conserved eukaryotic proteins, and consist of histone fold domains with N- and C-terminal tails extending from the

central histone fold. However, trypanosomal histones show extreme sequence divergence when compared to model eukaryotes, and even among different strains of the same species, like *T. brucei brucei* TREU 927 vs *T. brucei brucei* Lister 427. In addition to their role in DNA compaction, histones, and especially the N-terminal tails, are heavily modified by a wide range of PTMs that influence an assortment of regulatory functions (Kouzarides 2007). Comprehensive studies have demonstrated that histone PTMs can function collectively to form combinatorial PTM (cPTM) patterns that may accommodate more precise and rigorous epigenetic regulation, referred to as histone “cross-talk” (Su et al. 2014).

Investigation of *T. brucei* core histone PTMs revealed a lack of well-conserved modifications observed in other eukaryotes (Mandava et al. 2007; Janzen et al. 2006), and a number of PTMs were found to be unique to *T. brucei*. The N-terminal domain of histone H2A was found to be sparsely modified; A1 was observed in mono-methylated or acetylated forms, and acetylated K4 was only observed in the BF stage of the parasite (Fig 1.3, A). Mono-methylated A1 was also observed In *T. cruzi*, a closely related trypanosome that causes Chagas disease. In contrast, K4 was also found to be mono-methylated, but no acetylation was detected (Picchi et al. 2017). In contrast to *T. brucei*, the N-terminal tail of *T. cruzi* H2A was detected to harbor crotonylated K13, acetylated S19, and K21 present in either acetylated, tri-methylated, or 2-hydroxyisobutyrylated states. In contrast to the N-terminal of H2A, multiple PTMs have been detected on the C-terminal tail in *T. brucei*. Six acetylated lysines have been detected (on lysine residues 115, 119, 120, 122, 125, 125) of which

some (lysines 120, 122 and 128) correspond to conserved modifications observed and characterized in other eukaryotes (Mandava et al. 2007). Multiple PTMs were also observed on the H2A C-terminal tail of *T. cruzi*, with lysines 113, 116, 120, 121, 123 and 126 being acetylated. Of these K116, K120, K121, K123, and K126 seem to be functional equivalents of K115, K119, K120, K122, and K125 in *T. brucei.*

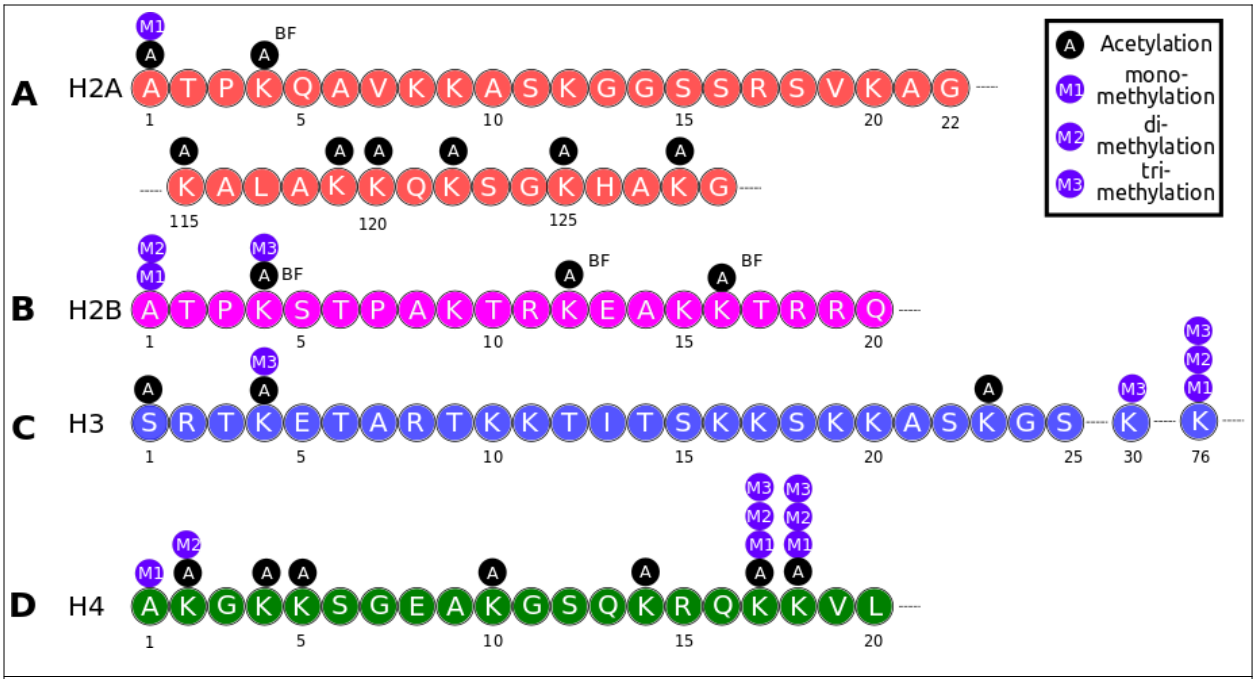


Figure 1.3: Post-translational modifications on *T. brucei* N-terminal histone tails.

Of the four trypanosomal core histones, H2B is the least conserved. PTM identification by mass spectrometric (MS) analyses and Edman degradation revealed a mere 5 PTMs on H2B, summarized in figure 1.3, B (Mandava et al. 2007; Figueiredo, Cross, and Janzen 2009). Histone H2B appeared scarcely modified with PTMs only observed on the N-terminal tail of H2B. Interestingly, three lysines at positions 4, 12, and 16 were found to be acetylated only in the BF stage of the parasite (Mandava et al. 2007).

Histone H3 is remarkably conserved across the eukaryotic lineage and displays a vast range of PTMs. Modification of the H3 N-terminal tail has been a subject of extensive study, with modifications on K4 and K9 demonstrated to have important epigenetic functions. Intriguingly, the universally conserved H3K9, associated with gene repression when tri-methylated, is totally absent in kinetoplasts (Lachner and Jenuwein 2002). Transcriptional repression of mouse olfactory gene clusters and var genes in *Plasmodium falciparum* is facilitated through heterochromatin formation via deposition of tri-methylated H3K9 (Cui and Miao 2010; Magklara et al. 2011). *T. brucei* employs an analogous epigenetic mechanism to maintain its vast VSG gene repertoire in a transcriptionally silent state. Although the *T. brucei* H3 tail contains a lysine 10 residue, it is unknown if the *T. brucei* H3K10 is the functional equivalent of H3K9 in other model eukaryotes (Fig. 1.4). No PTM has of yet been observed on the trypanosomal H3 K10 residue.

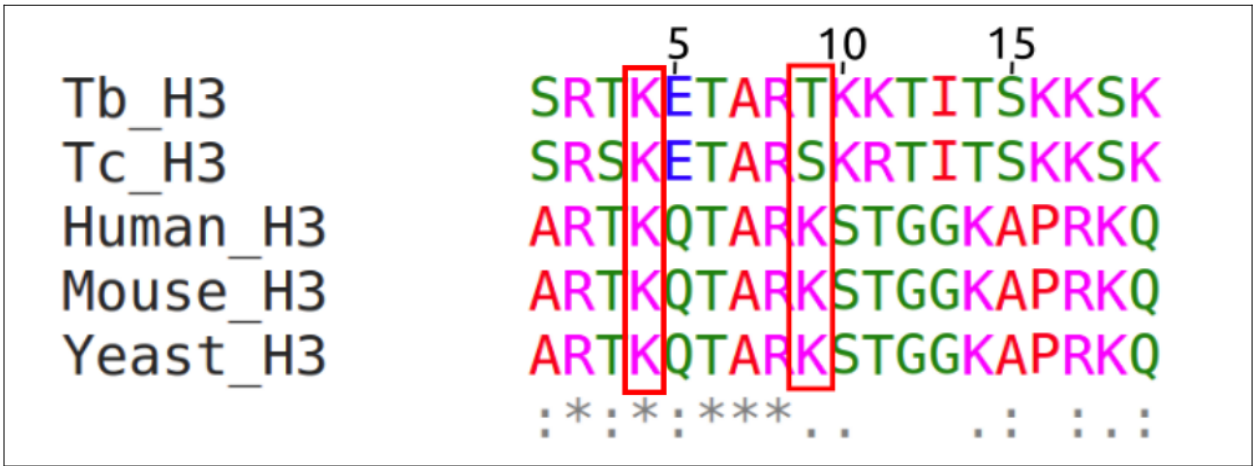


Figure 1.4. Alignment of the N-terminal tail segment of histone H3: *T. brucei* and *T. cruzi* N-terminal H3 tails do not contain the K9 residue (indicated in red block) like that seen in human, mouse and yeast H3. Conserved K4 residues are also indicated.

Histone H4 is the most conserved of the *T. brucei* core histones and contained the largest number of PTMs. All 7 lysine residues within the first 20 amino acids of the H4 N-terminal displayed acetylation marks (Fig. 1.3, D). Lysines 2, 17, and 18 were also found to be methylated to various degrees (Janzen et al. 2006; Mandava et al. 2007). Although only a handful of histone PTMs have to date been discovered, most of the identified marks are analogous to PTMs observed in *T. cruzi*. Recently, a comprehensive study of *T. cruzi* histone PTMs showed that trypanosomal histones are furnished with a multitude of PTMs present in intricate, combinatorial PTM patterns. In addition, the identification of a variety of PTMs on the *T. cruzi* histone variants strongly suggests that this is the case in *T. brucei*. Indeed, a bromodomain protein TbBDF2 has been demonstrated to selectively bind the hyperacetylated H2A.V N-terminal tail in *T. brucei*, evidence of cPTM employment in this trypanosome (Yang et al. 2017).

The *T. brucei* genome encodes a large repertoire of epigenetic regulators, including chromatin remodelers (like FACT, ISWI, TbSIR2RP, TDP1) (Navarro and Gull 2001; García-Salcedo et al. 2003; Denninger et al. 2010; Narayanan and Rudenko 2013), and histone modifiers and readers (such as HAT1 - 3, HDACs, DOT1A&B, TbHAT3, BDF3, etc.) to name but a few (Hake et al. 2006; Siegel et al. 2008; Kawahara et al. 2008; Figueiredo, Janzen, and Cross 2008). An excellent review by Figueiredo and colleagues provides a comprehensive overview of chromatin remodelers and histone readers, writers and erasers (Figueiredo, Cross, and Janzen 2009).

Over the past two decades, substantial evidence demonstrating the crucial role of epigenetic control in gene expression and transcriptional regulation in *T. brucei* has been generated. Despite the near complete absence of regulatory control at transcriptional level, like the constitutive transcription of pol II PTUs and lack of DNA regulatory sequences, *T. brucei* employs elaborate epigenetic mechanisms to regulate genome function. Co-localization of DNA modifications, histone variants, PTMs and chromatin binding proteins effectively demarcate the genome and define regions of transcription initiation and termination. In addition, life-cycle specific genes, like the pol I transcribed VSG in BF cells, are epigenetically regulated, and rely on a combination of epigenetic mechanisms to keep silent VSGs and BESs transitionally repressed.

It was previously believed that as a result of the early divergence of kinetoplasts, trypanosomes might possess a more compact epigenetic repertoire and, therefore, a less complex histone code. However, data strongly suggests that *T. brucei* does in fact possess elaborate epigenetic mechanisms that regulate nuclear processes including transcription, nucleosomal architecture, genome demarcation, and gene regulation. Although much progress has been made in elucidating the influence and effect of chromatin remodeling and nuclear architecture in *T. brucei*, from figure 1.3 it is strikingly obvious that research on the post-translational modification states of core histones has been sorely neglected. In the ten years since the last PTM survey of trypanosome histones, remarkable improvement in technology has allowed detailed analysis of protein post-translational modifications. Development of specialized techniques and

software now allow us to precisely identify, map, and quantitate combinatorial PTMs across entire proteins with unprecedented accuracy (Sidoli, Cheng, and Jensen 2012). Application of these techniques to investigate *T. brucei* histone modifications will surely uncover an abundance of PTMs and reveal cPTM patterns, as was the case in *T. cruzi*. Expansion on the understanding of histone epigenetics in such a divergent organism will undoubtedly shed light on parasite epigenetics, evolution of the epigenome and epigenetic processes, contribute to the understanding of the histone code, and uncover possible therapeutic epigenetic targets.

1.1) Aim of the study

This study aims to expand on the current knowledge of the *T. brucei* histone epigenetic repertoire by applying high-throughput middle-down MS techniques to elucidate the post-translational modification status of all canonical and variant core histones. In addition, the method allows investigation of combinatorial PTM patterns employed by this organism, providing further insight into the trypanosomal "histone code". Genomic distribution of selected (c)PTMs will be investigated via MNase-ChIP-seq to explore the possible functionality in genome demarcation and/or gene regulation and expression. Expansion on the current knowledge of *T. brucei* epigenetics, specifically histone PTMs and distribution, will assist in decoding and identifying epigenetic mechanisms that may be exploited for novel drug therapies.

2) Materials and Methods

All reagents used were of molecular biology grade purity.

2.1) Trypanosome strains and culture

In Vivo

For core histone extraction, bloodstream form *T. brucei brucei SmoxB4*, a cell line generated from the *T. brucei brucei* Lister 427 strain containing a T7 promoter sequence expressesing a Tet repressor (Poon et al. 2012), were cultured *in vivo* in rats. In brief, Swiss Albino rats were inoculated intraperitoneally with 500 µl phosphate buffered saline (PBS; 10 mM potassium phosphate; 100 mM sodium chloride, pH 7.4) containing 1 x 10^4 trypanosomes and monitored for parasitemia post-infection using the Hebert and Lumsden matching method (Herbert and Lumsden 1976). At peak parasitemia (10^8 parasites/ml blood), the animals were exsanguinated and blood harvested by cardiac puncture into heparinised vacutainers, yielding 8 - 10 ml blood per animal. Trypanosomes were purified using DEAE anion exchange chromatography (Lanham and Godfrey 1970). Trypanosomes were eluted with ice-cold PSG (PBS, 1% w/v glucose), centrifuged at 2,000 rcf for 10 min at 4˚C, yielding a pellet containing ~9 x 10^{10} parasites per sample, as estimated using a Neubauer hemocytometer. Cell pellets were then immediately stored at -80˚C until needed.

In Vitro

Bloodstream form Lister 427 (S16 221 Puro) *T. brucei* was cultured in HMI-9 medium (1L Isocove Modified Dubecco's medium (IMDM) powder (Sigma); 136 mg hypoxanthine; 182 mg L-cysteine; 110 mg sodium pyruvate; 38.7 mg thymidine; 28.2 mg bathocuproinedisulfonic acid; 3.024 g $NaHCO_3$, pH 7.4) supplemented with 15% v/v foetal calf serum (FCS), 0.1 mM β-mercaptoethanol, and appropriate drugs (0.2 µg/ml puromycin). Cells were cultured at 37ºC with 5% CO_2 as previously described (Hirumi and Hirumi 1989). Procyclic form (PF) *T. brucei* 427 cells were cultured anaerobically at 27ºC in SDM-79 medium (SDM-79, 2 g/L; $NaHCO_3$, pH 7.3), supplemented with 10% FCS and 5 µg/ml hemin. Table 2.1 summarizes the strains and appropriate drugs used in this study.

Histone isolation strains:			
Life Cycle	**Strain**	**Selectable marker**	**Drug concentration**
BF	*T. brucei* SmoxB4	n/a	n/a
PF	Amsterdam WT	n/a	n/a
MNase-ChIP-seq strains:			
Life Cycle	**Strain**	**Selectable marker**	**Drug concentration**
BF	S16 221 Puro	Puromycin	0.2 µg/ ml
PF	Amsterdam WT	n/a	n/a

Table 2.1. *T. brucei* strains used in this study.

2.2) Core histone isolation

Core histones were extracted using an acid extraction method specifically adapted for *T. brucei* (Nardelli, Ting, and Kim 2015). In brief, 1 x 10^{10} and 1 x 10^{9} cells/sample were harvested for BF and PF trypanosomes, respectively, washed with PBS (pH 7.4), and pelleted at 1,200 rcf for 5 min. Cells were resuspended in 1 ml Buffer A (0.25 M sucrose; 1 mM EDTA; 3 mM $CaCl_2$; 10 mM Tris-HCL pH 7.4; 0.5% saponin) by vortexing and centrifuged for 10 min at 4,000 rcf, 4˚C. The pellet was resuspended in 1 ml Buffer B (0.25 M sucrose; 1 mM EDTA; 3 mM $CaCl_2$; 10 mM Tris-HCL pH 7.4), and centrifuged for 10 min at 4,000 rcf, 4˚C. The resulting pellet was then resuspended in 1 ml Buffer C (1% Triton X-100; 0.15 M NaCl; 25 mM EDTA; 10 mM Tris-HCL pH 8.0), and centrifuged for 20 min at 12,000 rcf, 4˚C. The supernatant was removed and the chromatin pellet washed three times in 1 ml 100 mM Tris-HCL (pH 8.0) to remove any residual detergents, and then pelleted at 12,000 rcf, 4˚C for 5 min.

All reactions henceforth were carried out on ice unless otherwise stipulated. Protease inhibitors were added to all buffers just before use. Protease inhibitors used during core histone isolation included 0.1 mM PMSF, 10 mM sodium butyrate, protease inhibitor cocktail (cOmplete™ mini, EDTA-free, Roche) and phosphatase inhibitors (Halt™ Protease and Phosphatase Inhibitor Cocktail, Thermo Scientific) added to manufacturer's instructions.

The chromatin pellet was then resuspended in 600 µl of 0.4 N H_2SO_4. A disposable 1.5 ml Eppendorf dounce was used to dissociate histones from higher-order chromatin. The samples were incubated at 4˚C for 6 hours

under rotation. Acid-soluble proteins (contained in the supernatant) were recovered by centrifugation at 10,000 rcf for 15 min, 4˚C. The resulting supernatant was measured (~500 µl per sample), transferred to a fresh tube, and trichloroacetic acid (TCA) added to a final concentration of 33% v/v (~250 µl TCA). Samples were incubated overnight under gentle agitation at 4˚C, and pelleted at 16,000 rcf for 10 min, 4˚C. The resulting pellet was washed three times with ice-cold acetone and pelleted at 16,000 rcf for 5 min, 4˚C. The pellets were air dried for ~10 min and stored at -80˚C.

2.3) Proteolytic digestion

The acid extracted histone pellets were resuspended in 50 µl ddH_2O by sonication (5 min) followed by agitation for 1 hour at 4˚C. This was done four times for a total of four hours, and insoluble debris pelleted at 16,000 rcf for 10 min, 4˚C. Protein concentrations were determined by BCA method, using a Pierce™ BCA Protein Assay Kit (Thermo Scientific, Cat. no. 23225). For protein digestion and middle-down MS analysis, a total of 30 µg protein per sample was diluted with 75 mM ammonium acetate (pH 4.0) to a final concentration of 1 µg/µl. Proteolytic digestions were performed using Glu-C and Asp-N endoproteinases (Calbiochem) with an enzyme:sample ratio of 1:20 and 1:40, at room temperature and 37˚C, for 6 hours and overnight, respectively. Proteolytic digestion was assessed by SDS-PAGE analysis. Glu-C and Asp-N endoproteinases were chosen as they produce peptides ranging between 30 - 60 amino acids in length for all *T. brucei* histone isoforms, which are optimal for middle-

down MS analysis, and allows investigation of combinatorial PTMs on the N- and C-terminal tails of the histones.

2.4) Nano-Liquid chromatography

Histone peptides were separated using a nanoliter-flow Ultimate 3000 HPLC system (Thermo Scientific), as previously described (Sidoli et al. 2014). The nanoLC was equipped with a two-column set-up: a 5 cm precolumn (100 µm internal diameter) packed with C_{18} bulk material (ReproSil, Pur C18AQ 5 µm), and an 18 cm analytical column (75 µm id) with a pulled needle packed with PolycatA resin (3 µm particles, 1500 Å; PolyLC, Columbia, MD, USA). The buffer used for sample loading contained 0.1% (v/v) formic acid in ddH_2O. Buffers 1 and 2 were prepared as previously described (Young et al. 2009). Buffer 1 consisted of 75% acetonitrile (ACN) and 20 mM propionic acid (Fluka) adjusted to pH 6.0 with NH_4OH (Sigma-Aldrich). Buffer 2 consisted of 25% ACN adjusted to pH 2.5 with formic acid. The sample was loaded into the trap for 10 min with loading buffer at 5 µl/min. Peptides were separated using a gradient of 100% Buffer 1 (10 min), followed by 30-95% Buffer 2 (105 min) and 95-100% Buffer 2 (10 min) for column washing. The flow rate for the analysis was set to 230 nl/min.

2.5) Mass Spectrometric analysis

The nano-LC was coupled inline with an Orbitrap Fusion ETD (Thermo Scientific). A nano-electrospray ion source was used with spray voltage set at 2.4 kV, and capillary temperature set at 275°C. Data acquisition was performed in the Orbitrap instrument for both precursor and product

ions, with a mass resolution of 60,000 for MS and 30,000 for tandem MS (MS/MS). The MS acquisition window was set at 400 – 750 *m/z* with dynamic exclusion disabled. The precursor charges accepted for MS/MS fragmentation were 5–12, and isolation width set at 2 *m/z*. The six most intense ions with an MS signal higher than 5,000 counts were isolated for fragmentation using electron transfer dissociation (ETD) with an activation time of 20 ms. Three microscans were used for each MS/MS spectrum, and the AGC target was set to 2×10^5. The MS instrument was controlled by Xcalibur software (Thermo Scientific).

2.6) Middle-down MS data processing

Raw MS data files were processed with Proteome Discoverer (v1.4, Thermo Scientific), and spectra deconvoluted with the Xtract tool (Thermo Scientific) and searched using Mascot (v2.5, Matrix Science, London, U.K.). Mascot searches were carried out using the following parameters: MS peptide mass tolerance: 1.05 Da; MS/MS fragment mass tolerance: ± 0.02 Da; enzyme: Glu-C or Asp-N with 0 missed cleavages; Mass values: monoisotopic; Dynamic modifications: N-terminal acetylation, mono- and di-methylation (K,R), tri-methylation (K), and acetylation (K,S), propionylation (K), oxidation (F,P,S,T,V,W,Y), and phosphorylation (T,S,Y). To search the MS/MS spectra, canonical and variant histone protein sequences encoded in the *T. brucei* Lister 427 genome were retrieved from TriTryDB (release 25, December 2015) and used to construct a custom protein database, as outlined in table 2.2. This was done as the database available for MASCOT searches did not include the *T. brucei* 427 Lister strain, only the related *Trypanosoma cruzi* species.

Histone (canonical)	Gene ID	Histone (variant)	Gene ID
H2A	Tb427.07.2830	H2A.V	Tb427.07.6360
H2B	Tb427.10.10460	H2B.V	Tb427tmp.02.5250
H3	Tb427.01.2430	H3.V	Tb427.10.15350
H4	Tb427.05.4170	H4.V	Tb427.02.2670

Table 2.2. Gene ID's of histone sequences used for Mascot searches: Protein sequences for canonical and variant histones were retrieved from TriTryDB. Indicated are the gene IDs for the histone protein sequences used to construct a custom MASCOT searchable database.

Validation of assigned PTMs was done by exporting Mascot search results into a *.xml file, and processed with Histone Coder (Sidoli et al. 2014), using a tolerance of 30 ppm. Only PTMs with ≥1 site determining ion on either side of the mapped modification were accepted. Histone peptides were quantified using IsoScale Lite, an MS/MS-based PTM quantification software developed specifically for quantification of co-eluting, isobaric peptides, by using the total ion intensity from the MS^2 spectra to determine peptide abundance relative to the fragment ion ratio. IsoScale uses MASCOT result files and retrieves the total ion intensity of fragment ions from each MS/MS spectra for each peptide species, and differentiates abundance of isobaric/ co-fragmented peptides by calculating the fragment ion relative ratio of each peptide (Pesavento, Mizzen, and Kelleher 2006; Sidoli et al. 2014).

2.7) Calf thymus histone MS

To investigate whether the *T. brucei* H3 hydroxyproline 40 and H2A hydroxyproline 26 are artefactual PTMs caused by fortuitous oxidation of proline residues during the extraction procedure, or were genuine PTMs, we repeated the extraction procedure. BF *T. brucei* histones were acid extracted as described, and 100 µg calf thymus histones (CTH, Sigma) were spiked into the samples at the beginning of the protocol, or during the acid extraction step. The samples were analyzed as described above, and data searched using MASCOT for proline oxidation on histone H3 (Pro 40 in *T. brucei*, and Pro 43 in bovine) and histone H2A (Pro 26 in both *T. brucei* and bovine H2A). Analysis of the primary amino acid sequence of *T. brucei* histones H2A and H3 for putative proline hydroxylation sites was done using RF-Hydroxysite, an online bioinformatics tool that uses a random forest tree-based ensemble machine learning technique to identify putative hydroxyproline residues (Ismail, Newman, and KC 2016). A window size of 13 and 7 was selected using a cutoff threshold of 0.9 and 0.6 for H2A and H3, respectively.

2.8) Antibody generation

Custom polyclonal antibodies were raised in New Zealand white rabbits (GenScript Incorporated, Piscataway, USA) against specific histones, variants and PTMs (Table 2.3). In brief, synthetic peptides of ~15 amino acids in length were synthesized with a cysteine residue added to either the C- or N-terminus of the peptide to allow peptide conjugation to Keyhole Limpet Hemocyanin (KLH), which served as a carrier protein. Rabbits were immunised and test bleeds taken after the 3rd immunisation.

Antibodies were affinity purified and specificity determined using an indirect ELISA (GenScript Incorporated, Piscataway, USA). Antibodies designed to target a specific PTM were also verified for cross-reactivity against the unmodified version of the same peptide.

Antibody	**Peptide sequence**	**Modification**
anti-H3	K_{16}KSKKASKGSDAAS$_{29}$C	n/a
anti-H2A	CA_{70}AQQTKKTKRLTPR$_{83}$	n/a
anti-H2A.V	CE_{22}QASALTGGKLGGK$_{35}$	n/a
anti-H2A T2pho	$A_{1\ me1}$ *T_{2pho}*PKQAVKKASKC	Phosphorylated H2AT2
	$A_{1\ me1}$TPKQAVKKASKC	Unmodified peptide
anti-H2AK115ac	LN*K_{115ac}*ALAK*K_{120ac}*Q*K_{122ac}*SGKHAC	Hyperacetylated H2A K115acK120acK122ac
	LNKALAKKQKSGKHAC	Unmodified peptide
anti-H2AK125ac	LNKALAK*K_{120ac}*Q*K_{122ac}*SG*K_{125ac}*HAC	Hyperacetylated H2A K120acK122acK125ac
	LNKALAKKQKSGKHAC	Unmodified peptide
anti-H3K4me3	SRT*K_{4me3}*ETARTKKTITSKKC	Tri-methylated H3K4
	SRTKETARTKKTITSKKC	Unmodified peptide
anti-H3K10me3	SRTKETART*K_{10me3}*KTITSKKC	Tri-methylated H3K10
	SRTKETARTKKTITSKKC	Unmodified peptide
anti-H3HyPro40	CKTAQRRWR*$HyPro_{40}$*GTVA	H3 Hydroxyproline40
	CKTAQRRWRPGTVA	Unmodified peptide

Table 2.3: Synthetic peptides used for immunization. Modified amino acid residues are indicated in italic.

2.9) Antibody cross-reactivity testing

Serum was collected from test bleeds after the 3rd immunisation, affinity purified, and the antibody titer and cross-reactivity assessed via indirect ELISA (GenScript Incorporated, Piscataway, USA). Lyophilised antibodies and peptides were reconstituted in MilliQ H_2O to a final concentration of 1 mg/ml and 2 mg/ml, respectively. Antibodies from rabbits displaying the lowest levels of cross-reactivity were selected to manually assess cross-reactivity via a dot-blot Western using the Pierce™ Fast Western Blot Kit (Thermo Scientific, Cat. no. 35050). Five micrograms of both the modified and unmodified peptides were placed on a nitrocellulose membrane and air-dried for 10 min at room temperature to allow peptide fixation. The membrane was then probed with a 1:10 000 primary antibody dilution and visualised using a MyECL imager (Thermo Scientific).

2.10) Immunofluorescence microscopy

For immunofluorescence microscopy, 1 x 10^6 BF or PF cells were harvested by centrifugation (1,200 rcf, 10 min, room temperature), washed once in PSG (PBS, 1% w/v glucose) and fixed in 2% paraformaldehyde for 15 min. Cells were washed with PBS and settled on microscopy slides for 30 min at room temperature. Slides were then treated with 0.1% NP-40 for 5 minutes, washed twice with PBS for 5 min, and incubated with primary antibody in 1% bovine serum albumin (BSA; 1% w/v in PBS) for 1 hour in a humidity chamber.

Primary antibodies (anti-H2AK115ac and anti-H2AK125ac) were diluted 1:500 or 1:5000 in 1% BSA, respectively. Slides were washed three times for 5 min with PBS and incubated with goat-anti-rabbit Alexa594 secondary antibody (Thermo Scientific) at a 1:500 dilution in 1% BSA (1% w/v BSA in PBS) for 45 min in a humidity chamber. Slides were then washed three times for 5 min in PBS and mounted using Vectashield mounting medium with DAPI. Imaging was performed on a Zeiss Axioimager M1 with an AxioCam MRM camera using Axioimager software. For z-stacks, images were taken in 0.2 µm increments. Images were prepared using Image J software (Schneider, Rasband, and Eliceiri 2012).

2.11) MNase-ChIP-seq

T. brucei chromatin was digested using micrococcal nuclease (MNase), which preferentially digests unbound DNA yielding ~147bp nucleosome-bound DNA fragments, suitable for determining nucleosomal positioning via NGS sequencing and realignment. MNase digestion and DNA purification were performed as previously described (Nardelli, Ting, and Kim 2015). A native ChIP (N-ChIP) approach was used, i.e. no formaldehyde cross-linking, as cross-linking could bind other chromatin-associated proteins which may result in artefactual observations. In brief, 2.5 x 10^8 cells per sample were harvested by centrifugation (1,200 rcf, 5 min, room temperature), washed with PBS (pH 7.4) and pelleted at 1,200 rcf for 5 min. Cells were resuspended in 500 µl ice-cold cell suspension buffer (300 mM sucrose; 15 mM Tris–HCl pH 7.5; 5 mM $MgCl_2$; 15 mM NaCl; 60 mM KCl; 0.1 mM EDTA; 0.1 mM PMSF; and protease inhibitors (cOmplete™ mini, EDTA-free, Roche; Halt™ Protease and Phosphatase

Inhibitor Cocktail, Thermo Scientific). To this, 500 µl cell lysis buffer was added (0.4% NP-40; 1.0 mM Tris–HCl pH 7.5; 0.2 mM EDTA; 0.3 mM PMSF), homogenized by pipetting, incubated on ice for 5 min and then centrifuged at 4,000 rcf for 10 min at 4˚C.

The pellet was washed once with 1 ml ice-cold cell lysis buffer (without NP-40) and centrifuged again at 4,000 rcf for 10 min at 4°C. The nucleus enriched pellet was resuspended in 1 ml ice-cold MNase digestion buffer (0.32 M sucrose; 50 mM Tris–HCl pH 7.5; 4 mM $MgCl_2$; 1 mM $CaCl_2$; 0.1 mM PMSF and aforementioned protease inhibitors) and centrifuged at 4,000 rcf for 5 min at 4°C. The nuclei enriched pellet was resuspended with MNase digestion buffer to a final volume of 100 µl.

Chromatin was digested with 4U MNase (Worthington Biochemicals) at 37°C for 5 min. The reaction was quenched by addition of EDTA to a final concentration of 10 mM, and incubated on ice for 5 min. Digested chromatin was separated into two fractions (S1 and S2 - corresponding to less and more compacted chromatin states, respectively) by centrifugation (8,000 rcf for 5 min at 4°C) and the supernatant (S1 fraction) saved to a new tube. The pellet (S2 fraction) was resuspended in 400 µl lysis buffer and dialyzed overnight at 4°C in 1L lysis buffer under gentle agitation using a Pierce Slide-A-Lyzer Dialysis Cassette (3.5 K MWCO, 3 ml, Cat. no. 66330, Pierce). The sample was centrifuged at 500 rcf for 10 min at 4°C and the supernatant (S2 fraction) transferred to a new tube.

Fractions S1 and S2 were combined and diluted 10 times using ChIP dilution buffer (0.01% SDS; 1.1% Triton X-100; 1.2 mM EDTA; 16.7 mM Tris–HCl pH 8.1; 167 mM NaCl and protease inhibitors). From this, 200 µl was removed from each sample as input DNA.

To preclear the solution, 60 µl protein A coupled magnetic beads (Dynabeads, Protein A Cat. no. 100.02D, Invitrogen) was added to the solution and incubated for 30 min at 4°C under gentle agitation. The beads were separated using a magnetic rack and the supernatant transferred to a new tube. Next, 10 µl of a 1 µg/µl antibody solution was added to the supernatant and incubated overnight at 4°C under gentle agitation. Following immunoprecipitation, 80 µl Dynabeads were added to the solution and incubated for 2 h at 4°C under rotation to collect the antibody-protein-DNA complex. Dynabeads were pelleted using a magnetic rack, the supernatant aspirated, and washed three times each with 1 ml low-salt buffer (0.1% SDS; 1% Triton X-100; 2 mM EDTA; 20 mM Tris–HCl pH 8.1; 150 mM NaCl), followed by high-salt buffer (0.1% SDS; 1% Triton X-100; 2 mM EDTA; 20 mM Tris–HCl pH 8.1; 500 mM NaCl), LiCl wash buffer (0.25 M LiCl; 1% NP-40; 1% deoxycholate; 1 mM EDTA; 10 mM Tris–HCl pH 8.1) and finally six times with TE buffer (10 mM Tris–HCl pH 8; 1 mM EDTA). The DNA-protein complex was eluted twice with 200 µl elution buffer (1% SDS; 0.1 M $NaHCO_3$) for 15 min at room temperature and eluates combined. Input DNA samples were diluted with 10 mM Tris–HCl (pH 8) to a final volume of 400 µl.

To release the DNA from the histone octamer, 10 µl of 0.5 M EDTA and 15 µl of 1.5 M Tris–HCl (pH 6.5) were added to samples to which SDS and proteinase K were added to a final concentration of 0.5% (v/v) and 0.2 µg/µl, respectively, and incubated at 56˚C for 60 min. After proteolytic digestion, RNase A was added to a final concentration of 0.2 µg/µl and incubated at 37°C for 60 min. DNA was precipitated with phenol:chloroform:isoamylalcohol (25:24:1, v/v) and ethanol precipitated as follows: the volume of aforementioned eluent was measured, and, in this order, 1/10 vol. 3M NaAc (pH 5.2) and 3x vol. ice-cold 98% (v/v) EtOH added sequentially, vortexed, and the DNA precipitated at -20°C overnight. Precipitated DNA was pelleted at 4°C for 10 min at maximum velocity, washed with 70% (v/v) room temperature ethanol, and air dried. The DNA pellet was then resuspended in 15 µl TE buffer (10 mM Tris-HCl pH 8.0; 0.1 mM EDTA) and stored at -80°C.

2.12) Paired-End DNA sequencing

DNA repair, adapter ligation, and library preparation were performed using the Illumina Truseq Chip-seq kit (ARC-BTP Onderstepoort). DNA was sequenced with an Illumina HiSeq 2500 platform using one flow cell, producing 2x 125 nt sequences. All reactions were carried out to manufacturer's instructions. Regrettably, QC analysis of sequenced DNA revealed that some of the ChIP samples were of insufficient quality or quantity, and were therefore discarded from further analysis, including BF and PF histone H3 technical replicates, BF and PF H2A, H2A.V and H2AT2 phosphorylation samples, and BF and PF H2A K115ac technical replicates.

2.13) Bioinformatic data analysis

The 125 nt paired-end reads obtained was realigned to the *T. brucei* Lister 427 genome (build 37, http://tritrypdb.org/tritrypdb/) using a custom-built workflow analysis (Fig 2.1) on the Galaxy bioinformatics platform using default settings, unless otherwise stipulated (Afgan et al. 2016).

In brief, *Trim_Galore!* was used for quality trimming and adapter clipping, as well as quality control, using default settings with minimum read length set to 115 bp and maximum read length to 125 bp. *FastQC* was used to assess the sequence quality after trimming (Andrews 2010). The trimmed reads were re-aligned to the *T. brucei* Lister 427 genome with *Bowtie 2* (Langmead et al. 2009), using default settings. This was done for the input DNA (control) as well as the immunoprecipitated (treatment) DNA. *Bowtie 2* *.bam alignment files were analysed for read enrichment using *MACS2* (Zhang et al. 2008), taking the input DNA as 'control', and ChIP DNA as 'treatment' files. The modified parameters used for peak calling included a user-defined genome size (26,75 Mb), 200 bandwidth for fragment size computation (used in building the shifting model), a minimum false discovery rate (FDR) of 0.01, composite broad regions selected, and duplicate tags allowed set to 'auto'. *Bowtie 2* *.bam output files, as well as *MACS2* bedgraph output files, were inspected manually using integrated genome viewer (Robinson et al. 2011; Thorvaldsdóttir, Robinson, and Mesirov 2013) to visually inspect sequence realignment data and peaks called by *MACS2 to* check for inconsistencies as well as alignment artefacts. *MACS2* called peaks data files can be accessed in *.bed format at: https://usegalaxy.eu/u/jpm/h/macs-2-called-peaks---jp-maree or at https://drive.google.com/open?id=1zHxZVPlVqBUh3_D2onmVXP2pTYO_8H7N

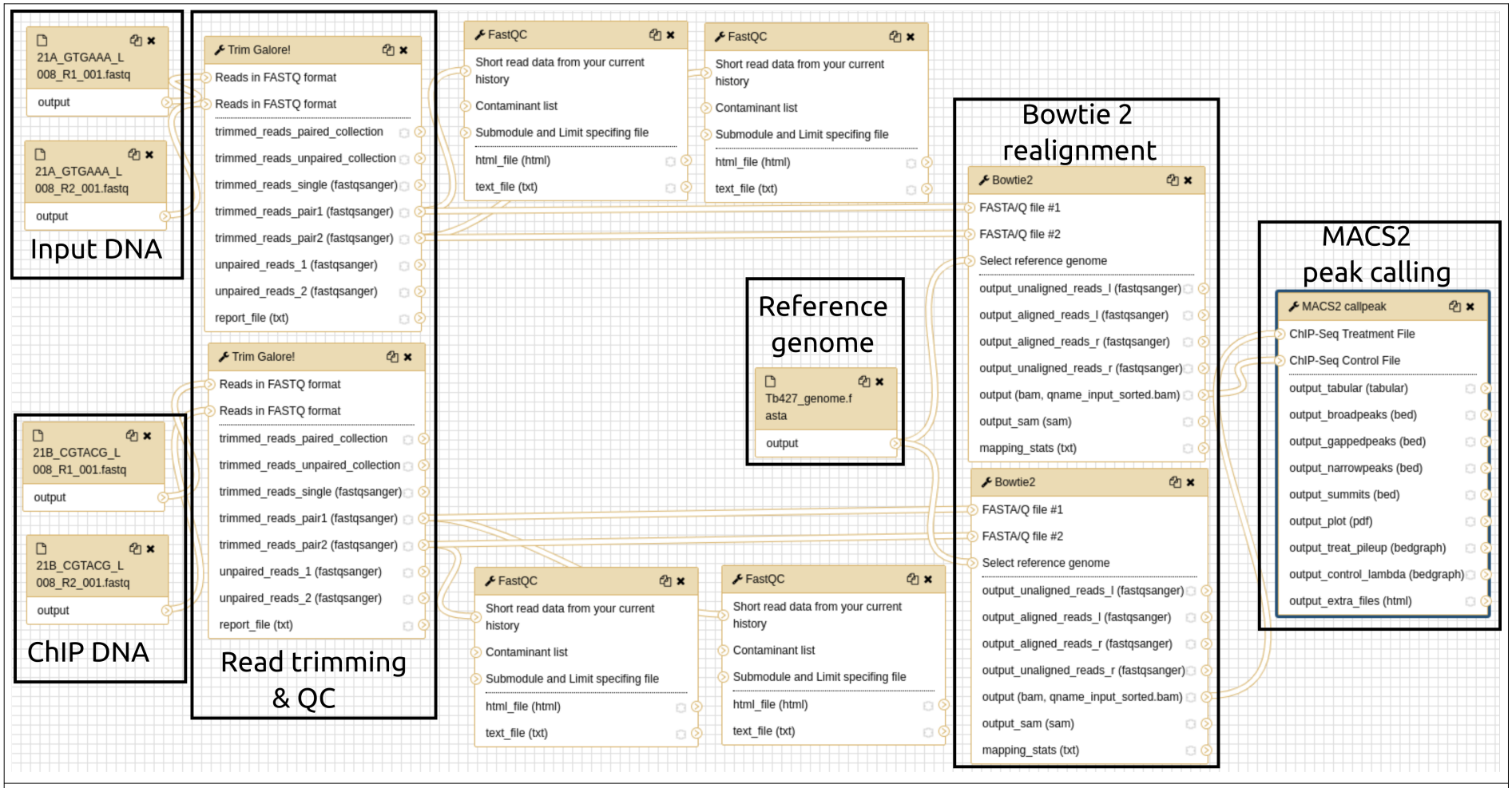


Figure 2.1. Custom workflow employed on the Galaxy platform: Input and ChIP sequences were trimmed and read quality checked before realignment with *Bowtie 2*, followed by peak calling by *MACS2*.

2.14) Genomic regions intersect

Input files were obtained from TriTrypDB (v.37) or assembled from previously published work and curated manually. Data sets obtained from the *T. brucei* 927 strain were mapped back to the *T. brucei* 427 strain and curated manually. BAM, GFF, GFF3, and GTF files were generated and manually curated before the *intersect intervals* tool was evoked using Galaxy. Internal stop/start sites (ISS) were compiled from previous data (Siegel et al. 2009; Reynolds et al. 2016; Wedel et al. 2017) and manually curated.

Dataset	Obtained from	Remapped (927 to 427)
dSSR	TriTripDB, Maree et al. 2017	Yes[1,2]
Internal stop/start	TriTripDB, Wedel et al. 2017	Yes[1,3]
cSSR	TriTripDB	Yes[1]
SAS	Siegel et al. 2009	Yes[1,2,3]
SL RNA	TriTripDB	No
rRNAs	TriTripDB	No
snoRNA	TriTripDB	No
tRNAs	TriTripDB	No
VSG	TriTripDB, Maree et al. 2017	No
Subtelomeric	TriTripDB, Maree et al. 2017	No
Procyclin	TriTripDB	No
ORC	Tiengwe et al. 2012	Yes[2]
Centromeres	Echeverry et al. 2012)	Yes[2]

Table 2.4. Summary of genomic regions used to examine localization of PTMs in the *T. brucei* 427 genome:

[1] *T. brucei* 427 genome was perused and SSRs mapped manually.

[2] Positions identified and mapped in a previous study (Maree et al. 2017).

[3] From previous data (Siegel et al. 2009; Reynolds et al. 2016; Wedel et al. 2017)

3) Mass spectrometric analysis of core and variant histone cPTM states

3.1) Isolation of core histones from BF and PF *T. brucei*

Since their discovery in 1884 by Albrecht Kossel (Kossel A 1884), histones have been extensively studied using acid-based extraction protocols, typically by using HCl or H_2SO_4 acids. This technique allows isolation and enrichment of histone proteins with minimal contamination from other major nuclear molecules, like RNA, DNA and non-histone proteins (Shechter et al. 2007).

Ordinarily, acid extraction of histones is done using purified nuclei, which can be achieved successfully in bulk from chicken erythrocytes (Neelin et al. 1964). However, as it is technically extremely challenging to purify whole, intact nuclei from *T. brucei*, an adapted protocol was used where the cellular wall is permeabilised, and the cytosolic content removed by consecutive wash steps, yielding trypanosome "ghosts" - a porous cellular membrane containing the cell nucleus and some other organelles (Crowe et al. 1984). This results in significantly less contamination by non-nuclear proteins. *T. brucei* histones were acid extracted as described (Nardelli, Ting, and Kim 2015) using 0.4 N HCl, resulting in the enrichment of all canonical and variant histones (Fig 3.1). Eight samples were generated for both BF and PF trypanosomes, yielding ~143 µg total protein per sample (2.3 - 2.9 µg/µl) for BFs, and ~83 µg (1.5-1.8 µg/µl) total protein per sample for PFs, as determined by BCA analysis.

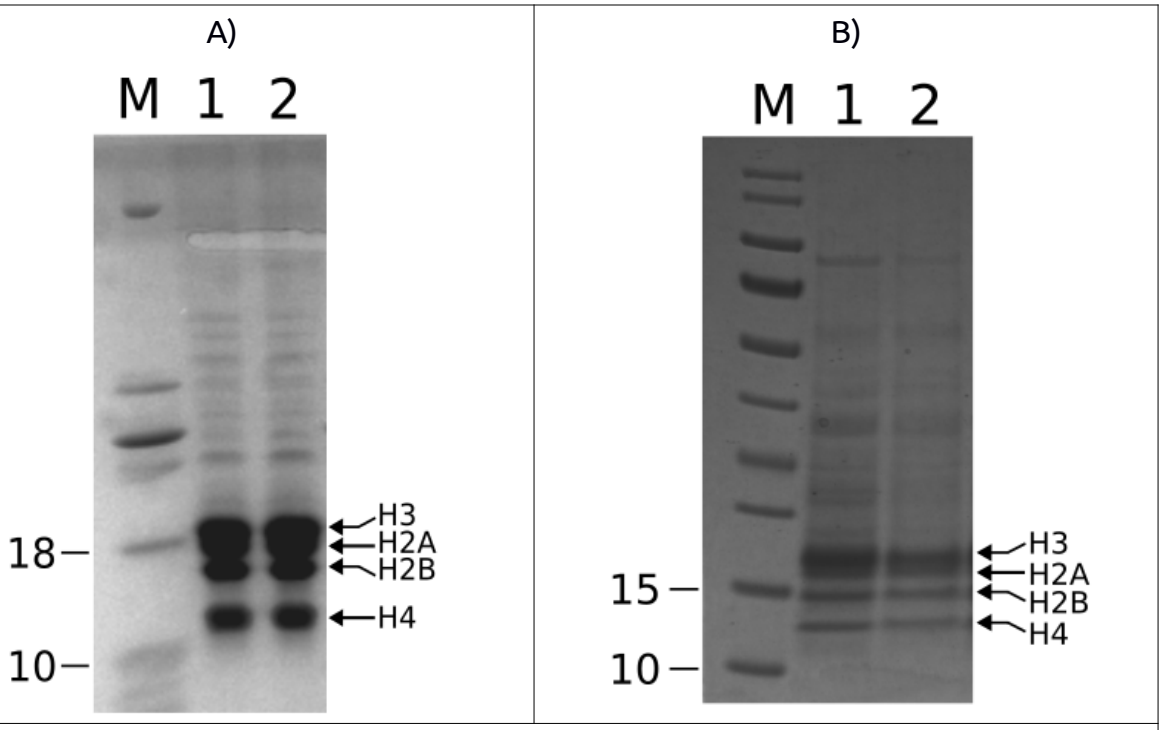


Figure 3.1. SDS-PAGE of bulk histones isolated from BF and PF trypanosomes: (**A**) 15% SDS-PAGE of bulk histones from BF trypanosomes, (**B**) 12% SDS-PAGE of bulk histones from PF trypanosomes. M = Marker, samples from histone extractions were loaded in duplicate, numbered 1 and 2 in both **A** and **B**. Sizes of molecular weight markers are indicated on the left-hand side (in kDa).

3.2) Peptide generation by proteolytic digestion with Glu-C or Asp-N

It was previously shown that instead of functioning individually, complex patterns of histone post-translational modifications are employed by a cell to define functional chromatin states, the "histone code" (Strahl and Allis 2000; Su and Denu 2016).

Recently, the existence of combinatorial histone PTMs have been observed in a related trypanosome, *T. cruzi*, on both canonical and variant histones (Picchi et al. 2017). Although several histone PTMs have been observed on canonical *T. brucei* histones (Mandava et al. 2007), and the use of histone variant deposition in genome demarcation investigated (Siegel et al. 2009; Reynolds et al. 2016), no study to date has examined the occurrence of combinatorial PTMs (cPTMs), or the PTM state of histone variants in this trypanosome. To investigate and identify combinatorial histone PTM patterns on canonical and variant histones in *T. brucei*, bulk histones were digested with either endoproteinase Glu-C or Asp-N, which preferentially cleaves peptides C-terminal to glutamic acid or N-terminal to aspartic acid residues, respectively. These endo-proteinases were chosen as they yield peptide fragments of 30 - 60 amino acids in length for all histone isoforms, amenable to middle-down MS analysis. Trypsin digestion would generate a multitude of very short peptides which is intractable to informative MS/MS analysis. *In silico* digestions were performed using either Glu-C or Asp-N endo-proteinases (Supplementary document R1) using the ExPASY PeptideCutter utility (Gasteiger et al. 2005). Bulk histones were digested and the efficiency of the digestions examined by gel electrophoresis (Fig 3.2). Digestion by Glu-C or Asp-N endo-proteinases resulted in adequate protein fragmentation, yielding peptide fragments amenable to middle-down MS investigation of the combinatorial PTM patterns.

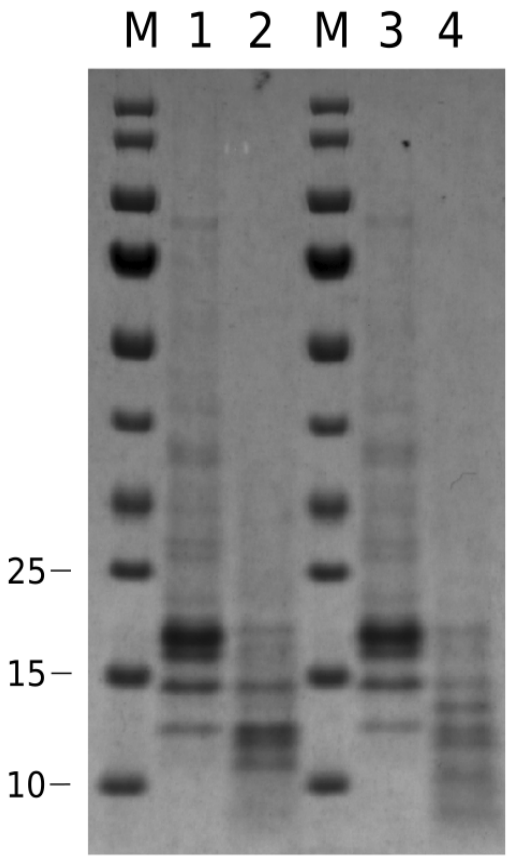


Figure 3.2. Analysis of bulk histone digestions by Glu-C and Asp-N using a 12% SDS-PAGE: M = PAGE-Ruler marker, 1 = undigested sample, 2 = Glu-C digested sample, 3 = undigested sample, 4 = Asp-N digested sample.

3.3) LC-MS/MS analysis of histone modification states

To avoid the necessity of large amounts of bulk histones needed to isolate individual core histone isoforms via conventional C_{18} reverse-phase high-performance liquid chromatography (RP-HPLC), we utilized a hybrid LC/LC setup. This setup negates the need for RP-HPLC histone isolation before proteolytic digestion and subsequent mass-spectrometric analysis. This method employs an RP trap column (packed with C_{18} bulk material) coupled to a WCX-HILIC column, allowing samples to be solubilized and loaded in an aqueous solution. The WCX-HILIC column, packed with PolycatA resin - a weak cation exchange resin, can selectively separate peptide charge variants, where the extensively modified peptides have altered retention times relative to less modified peptides, as peptide modification will alter amino acid charge states (such as negation of the positive charge of the basic lysine residues by acetylation) thereby altering total peptide charge, resin interaction, and retention time. Therefore, peptide retention time adds an additional factor which aids with differentiating between isobaric peptide species. Figure 3.3 shows the total ion chromatogram (TIC) for Glu-C digested histones, and the extracted ion chromatogram for the unmodified N-terminal histone H3 peptide. The composite LC/LC setup was connected inline to an Orbitrap Fusion ETD high-resolution mass spectrometer, allowing analysis of bulk, crude histone samples, while requiring up to 10-fold less starting material (Sidoli et al. 2014). Figure 3.4 shows the LC-MS/MS workflow output; WCX-HILIC chromatogram, MS^1, and MS^2 spectra of a 42 amino acid peptide originating from the H3 N-terminal following Glu-C digestion.

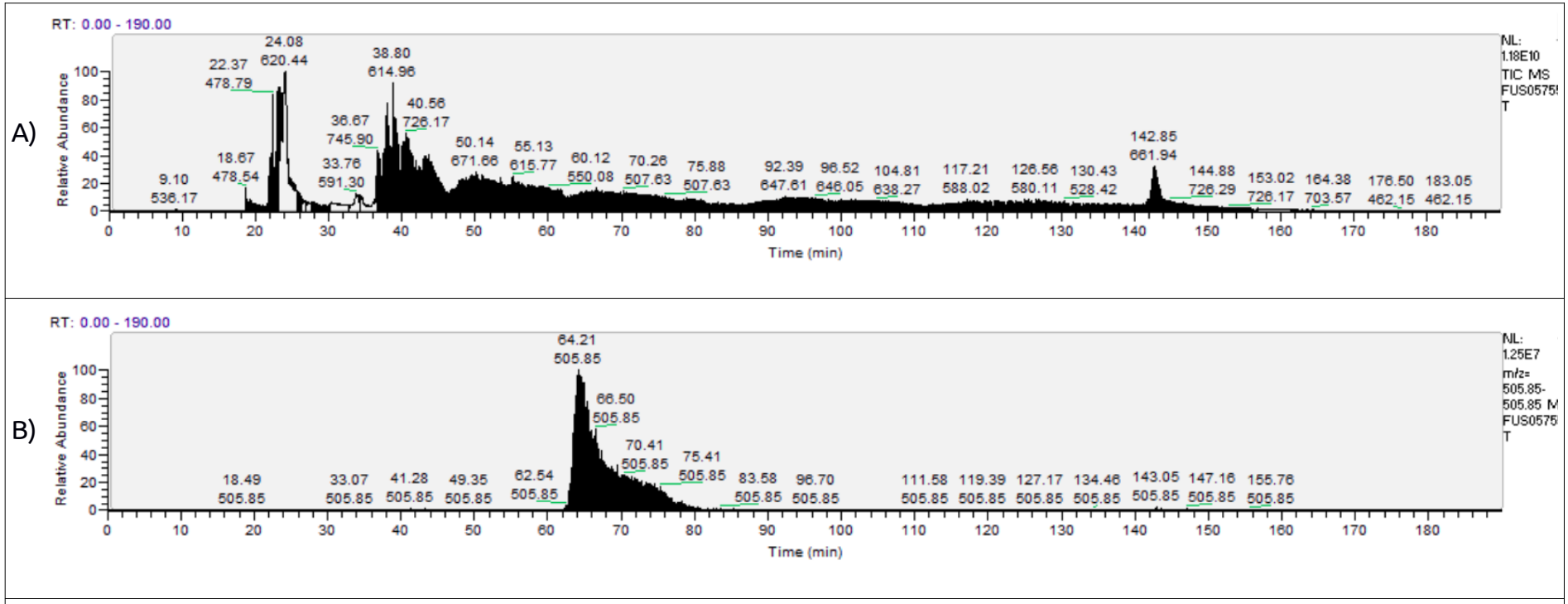


Figure 3.3. WCX-HILIC chromatograms: (**A**) Total ion chromatogram showing elution profile of bulk histones from bloodstream form trypanosomes, digested with endo-proteinase Glu-C. (**B**) Extracted Ion Chromatogram (XIC) showing the unmodified N-terminal histone H3 peptide (residues 6 - 47)

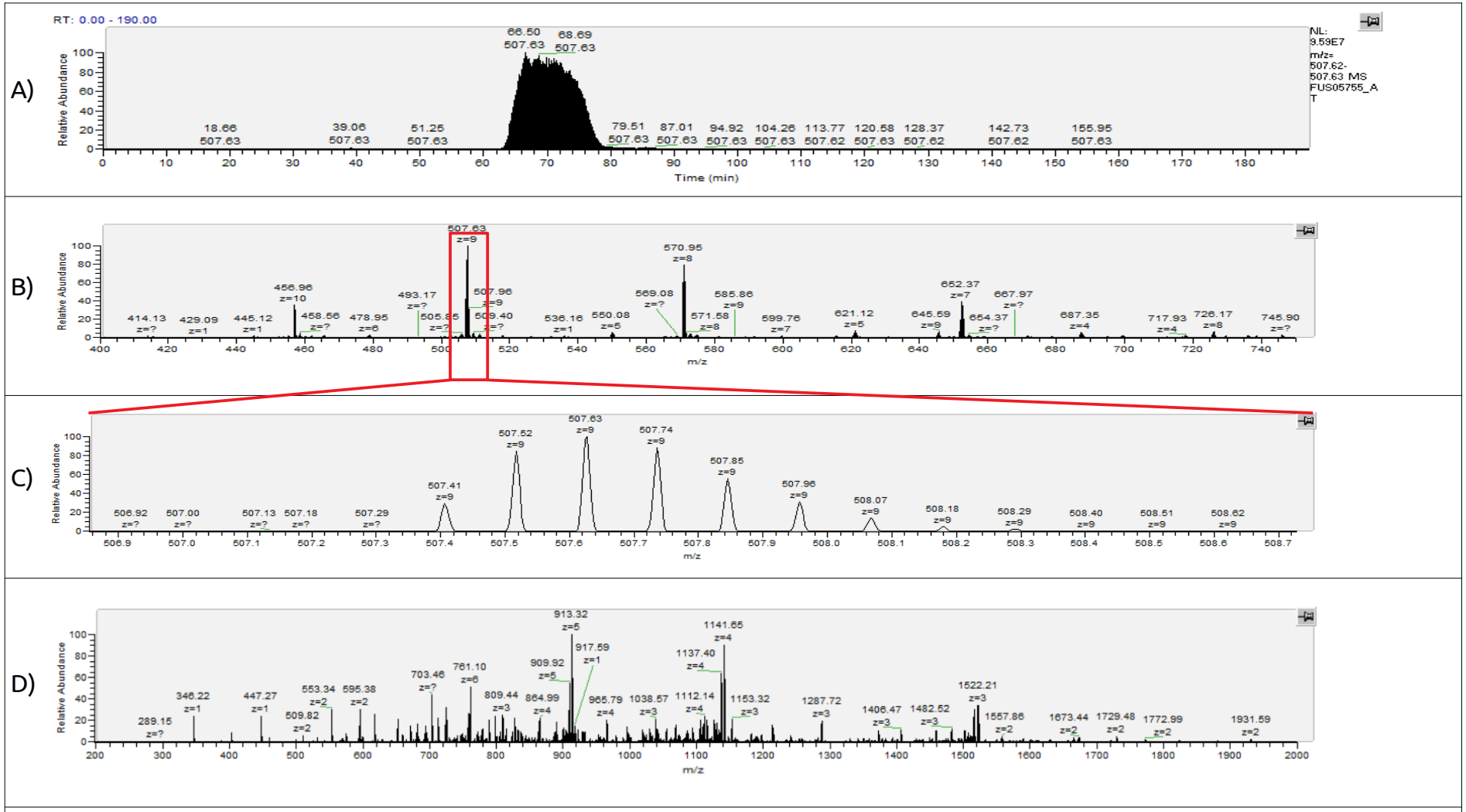


Figure 3.4: (**A**) WCX-HILIC chromatogram of a hydroxylated histone H3 peptide. (**B**) MS^1 scan of the same peptide (mass resolution: 60,000) showing different charge states of the peptide. (**C**) Isotopic distribution of peptide modifications. (**D**) MS^2 scan (mass resolution: 30,000).

Processed and deconvoluted MS^2 spectra were searched using Mascot (v2.5, Matrix Science, London, U.K.) with the following parameters:

MS^1 peptide mass tolerance: 1.05 Da

MS^2 fragment mass tolerance: ± 0.02 Da

Enzyme: Glu-C or Asp-N with 0 missed cleavages

Mass values: monoisotopic

Dynamic modifications:

- N-terminal acetylation
- mono- and di-methylation (K, R)
- tri-methylation (K)
- acetylation (K, S)
- propionylation (K)
- oxidation (F, P, S, T, V, W, Y)
- phosphorylation (T, S, Y)

The available Mascot database did not include protein sequences from the *T. brucei* Lister 427 strain, only the related kinetoplastid *T. cruzi.* Although histones are among the most conserved proteins, trypanosomes diverged very early on from the main eukaryotic lineage (He, Fiz-Palacios, Fu, Tsai, et al. 2014; Cavalier-Smith 2010) and display a remarkable degree of histone sequence divergence compared to higher eukaryotes, even among closely related trypanosome species (see Table 3.1 and Supplementary document R2).

Histone	Percent Identity
H2A	84.96%
H2B	81.08%
H3	92.42%
H4	96.97%

Table 3.1. Percentage identity between *T. brucei* (Lister 427) and *T. cruzi* (CL Brenner) core histones: Percentage identity was computed using Clustal Omega (Sievers et al. 2011).

Some amino acid substitutions can mistakenly be identified as modifications based on the observed mass shifts (like D›E [14.02 *m/z* shift], V›I [14.02 *m/z* shift], and S›T [14.02 *m/z* shift] as mono-methylation [14.02 *m/z* shift]; A›V [28.03 *m/z* shift], K›R [28.01 *m/z* shift] as di-methylation [28.03 *m/z* shift]; S›E [42.01 *m/z* shift] as acetylation [42.01 *m/z* shift]; etc.) and an artefactual PTM mistakenly assigned to an unmodified residue (Kim, Zhong, and Pandey 2016). Therefore, the need arose to create a custom-made database containing canonical and variant histone sequences from the current *T. brucei* genome to search and correctly identify modified residues (Supplementary document R3). Identification and assignment of histone PTMs proved to be challenging for histones H4 and H4.V, as they share 94.9% similarity and 89.9% identity (Fig. 3.5), with some amino acid substitutions resulting in artefactual PTM identification. However, identified PTMs on H4 and H4.V that passed validation by Histone Coder (discussed below) was recorded.

Figure 3.5. Sequence alignment of *T. brucei L*ister 427 histones H4 and H4.V, showing amino acid substitutions.

Assigned PTMs were validated by processing Mascot search results with Histone Coder (Sidoli et al. 2014), a software specifically developed to unambiguously determine the localization of modifications assigned by Mascot through counting the number of MS/MS ions in a given spectrum.

Glu-C cleavage of histone H2A resulted in two large peptide fragments, containing residues 1 - 60 (N-terminal peptide) and 65 to 133 (C-terminal peptide). The N-terminal peptide of H2A was found to be sparsely modified. About ~65% of H2A is N-terminally mono-methylated on alanine 1 in both life cycles. The proline 26 residue was found to bear a 16 Dalton mass shift, indicating that this residue may be hydroxylated (HyPro26, discussed in section 3.4). H2A N-terminal modifications are summarised in table 3.2. PTM quantitation was performed using IsoScale Lite (Sidoli et al. 2014). Interestingly, two combinatorial modification patterns were observed for this peptide, one bearing phosphorylated threonine 2, the other hydroxylated proline 26, with both containing

mono-methylated alanine 1. These combinatorial PTMs may represent differential modification patterns used to mark different loci, or recruit different chromatin modifying enzymes. Modification status of the N- and C-terminal H2A peptides were quantitated as these modifications were of interest to the current study (Tables 3.2 and 3.11, respectively).

Modification	**BF1**	**BF2**	**PF1**	**PF2**
Unmodified	8.1 %	7.5 %	8.3 %	7.6 %
*A_{1me1}	65.8 %	59.9 %	66.8 %	67.8 %
A_{1ac}	8.7 %	14.9 %	10.1 %	10.6 %
$A_{1me1}T_{2phos}$	9.5 %	10.8 %	10.1 %	8.7 %
$A_{1me1}HyPro_{26}$	7.8 %	7.0 %	4.7 %	5.3 %

Table 3.2. Quantitation of the H2A N-terminal peptide for two BF and two PF samples.

* Analogous to PTMs observed in *T. cruzi (*Picchi et al. 2017).

Tables 3.3 to 3.10 summarizes PTMs identified by Mascot and assigned by Histone Coder for BF and PF samples. Columns show Mascot score, Mascot expectation value ($p<0.05$), and the assigned histone code in standard Brno PTM nomenclature (Zhao and Garcia 2015). Site-determining ions, such as |2-K4me1-6| are interpreted as follows: mono-methylated lysine 4 has 2 site determining ions to the left {A1:2/6(z+1:27;z+2:27;)} and six to the right {K8:6/8(c:4;c:7;z+1:21;z+1:22;z+2:23;z+1:24;)}, unambiguously placing mono-methylation at lysine 4.

Mascot Score	Expectation Value	Histone code	Site Determining Ions
60.69	8.6E-07	T2phoK4me1	|1-T2ph-8||1-K4me1-4|
55.84	5.70E-05	*K4me1	|2-K4me1-2|
68.82	1.50E-06	K4ac	|2-K4ac-2|
44.28	8.2E-4	K9me1	|1-K9me1-5|
59.58	3.5E-4	S11ac	|2-S11ac-3|
24.47	3.9E-3	K115me1	|4-K115me1-7|
44.52	3.50E-05	*K115ac	|2-K115ac-1|
21.4	7.2E-3	K119me3	|8-K119me3-3|
68.06	1.60E-07	*K119ac	|8-K119ac-3|
25.54	2.8E-4	K120me3	|3-K120me3-4|
80.49	8.90E-09	*K120ac	|3-K120ac-4|
61.21	7.60E-07	K122ac	|1-K122ac-4|
62.32	2.20E-03	S123ac	|1-S123ac-2|
37.31	1.9E-4	K128ac	|1-K128ac-39|
43.14	8.80E-04	K115me3K120ac	|22-K115me3-9||2-K120ac-6|
45.97	3.00E-04	K115me3K120ac	|14-K115me3-6||2-K120ac-4|
38.37	0.001	K115me3K119acK120ac	|20-K115me3-8||8-K119ac-3||3-K120ac-5|
41.13	5.40E-04	K115acK119acK120ac	|1-K115ac-5||5-K119ac-2||2-K120ac-1|
49.41	2.10E-04	K115acK120me3	|3-K115ac-9||3-K120me3-4|
109.83	1.90E-10	K115acK120ac	|2-K115ac-9||3-K120ac-2|
99.2	2.20E-09	K115acK122ac	|2-K115ac-9||1-K122ac-2|
64.48	6.50E-06	K115acK125ac	|2-K115ac-9||1-K125ac-4|
75.73	4.90E-07	K115acK128ac	|2-K115ac-9||1-K128ac-1|
71.49	1.30E-06	*K119acK120ac	|1-K119ac-3||3-K120ac-3|

Table 3.3. PTMs assigned to histone H2A. * Analogous to PTMs observed in *T. cruzi.*

Mascot Score	Expectation Value	Histone code	Site Determining Ions
22.27	0.02	*K55ac	|5-K55ac-1|
45.19	7.90E-05	*K59ac	|4-K59ac-35

Table 3.4. PTMs assigned to histone H2A.V. * Analogous to PTMs observed in *T. cruzi.*

Mascot Score	Expectation Value	Histone code	Site Determining Ions
53.11	9.7E-4	K4me1	|1-K4me1-1|
42.87	0.01	K9me1	|1-K9me1-3|

Table 3.5. PTMs assigned to histone H2B.

Mascot Score	Expectation Value	Histone code	Site Determining Ions
80.01	1E-08	K4me3	|1-k4me3-9|
39.99	0.0057	*K4acK7acK15acK17ac	|1-K4ac-2||2-K7ac-5||5-K15ac-1||1-K17ac-2|
39.88	0.0058	K4acK7acK15acK17me3	|1-K4ac-2||2-K7ac-4||4-K15ac-2||2-K17me3-1|
49.06	0.0007	K4acK7acK15acK19me3	|1-K4ac-2||2-K7ac-4||4-K15ac-2||2-K19me3-7|
18.95	0.022	K7acR8me2	|1-K7ac-2||1-R8me2-1|
52.27	5.9E-05	K7acK15acK19me3	|2-K7ac-5||5-K15ac-1||2-K19me3-2|
48.94	0.00013	K7acK15acK27me3	|2-K7ac-5||5-K15ac-1||1-K27me3-6|
47.62	0.00017	K7acK17me3K19me3	|2-K7ac-5||1-K17me3-2||2-K19me3-2|
61.8	3.7E-05	*K7acK15acK17acK19ac	|2-K7ac-4||4-K15ac-3||3-K17ac-2||2-K19ac-9|
48.84	0.00074	K7acK15acK17acK19me3	|2-K7ac-4||4-K15ac-3||3-K17ac-2||2-K19me3-9|
48.84	0.00074	K7acK15acK17acK19me1R20me2	|2-K7ac-4||4-K15ac-3||3-K17ac-2||2-K19me1-1||1-R20me2-8|
26.75	0.041	R8me2K19ac	|1-R8me2-2||3-K19ac-8|
38.67	0.0035	*K15acK17acK19ac	|2-K15ac-3||3-K17ac-2||2-K19ac-7|
31.33	0.019	K15acK17acK19me3	|2-K15ac-3||3-K17ac-2||2-K19me3-7|
34.78	0.0086	K17acK19ac	|1-K17ac-1||1-K19ac-6|

Table 3.6. PTMs assigned to histone H2B.V. * Analogous to PTMs observed in *T. cruzi.*

Mascot Score	Expectation Value	Histone code	Site Determining Ions
80.03	9.9E-09	K4ac	|1-K4ac-9|
113.15	4.8E-12	K4me3	|2-K4me3-16|
57.11	0.0012	*K23me1	|5-K23me1-3|
68.73	8.4E-05	K32me1	|7-K32me1-1|
80.03	6.2E-06	R36me1	|6-R36me1-2|
59.23	0.00013	R66me1	|3-R66me1-3|
54.4	0.00039	R69me1	|1-R69me1-16|
67.36	4.4E-05	S72ac	|15-S72ac-3|
87.93	3.8E-07	*K76ac	|2-K76ac-21|
125.38	6.9E-11	*K76me3	|26-K76me3-53|
88.97	3E-07	*R66me1R69me2	|10-R66me1-1||1-R69me2-2|
72.27	1.4E-05	R66me1K76me2	|12-R66me1-1||1-K76me2-38|
84.49	8.4E-07	R66me2K76me1	|9-R66me2-1||2-K76me1-38|
68.7	3.2E-05	R69me1K76me2	|5-R69me1-3||3-K76me2-45|
90.02	2.4E-07	R69me2K76me1	|7-R69me2-2||2-K76me1-44|
40.31	0.034	K120acR132me2	|4-K120ac-31||1-R132me2-32|

Table 3.7. PTMs assigned to histone H3. * Analogous to PTMs observed in *T. cruzi.*

Mascot Score	Expectation Value	Histone code	Site Determining Ions
27.8	0.018	K82ac	|1-K82ac-15|
51.71	7.2E-05	K82me3	|22-K82me3-22|

Table 3.8. PTMs assigned to histone H3.V. * Analogous to PTMs observed in *T. cruzi.*

Mascot score	Expectation Value	Histone code	Site Determining Ions
37.35	0.00018	K2me1	|1-K2me1-1|
54.27	0.00082	*K2me3	|1-K2me3-3|
73.33	1.00E-05	*K2ac	|1-K2ac-3|
53.97	0.00088	K4me3	|1-K4me3-1|
116.5	7.00E-10	*K4ac	|3-K4ac-2|
109.86	9.60E-10	K5me1	|1-K5me1-5|
57.37	0.00034	*K5ac	|1-K5ac-4|
70.09	3.10E-05	*K10me3	|1-K10me3-7|
73.92	1.30E-05	*K10ac	|1-K10ac-3|
54.09	0.00088	K14me1	|1-K14me1-2|
61.23	0.00014	K14me3	|7-K14me3-3|
52.72	0.0012	*K14ac	|2-K14ac-6|
87.93	3.00E-07	K17me3	|6-K17me3-1|
59.88	0.00051	K18me1	|1-K18me1-1|
114.04	7.70E-10	K18me2	|2-K18me2-6|
97.61	3.20E-08	K18me3	|2-K18me3-32|
Table 3.9. PTMs assigned to histone H4. * Analogous to PTMs observed in *T. cruzi.*			

Mascot Score	Expectation Value	Histone code	Site Determining Ions
64.16	0.00043	R21me1	|4-R21me1-5|
68.03	1.60E-07	R29me1	|1-R29me1-8|
32.46	0.00057	R33me1	|1-R33me1-3|
Table 3.10. PTMs assigned to histone H4.V.			

The C-terminal tail of histone H2A has previously been shown to contain multiple acetylated lysine residues between K115 and K128 (Mandava et al. 2007). MS analysis confirmed that the *T. brucei* H2A C-terminal tail was hyperacetylated, with up to 5 acetylated lysines co-occurring (Table 3.11).

Modification	**BF1**	**BF2**	**PF1**	**PF1**
Unmodified	48.5 %	40.0 %	46.9 %	44.5 %
Ac x 1	14.3 %	12.7 %	5.3 %	5.3 %
Ac x 2	6.9 %	7.0 %	4.8 %	5.6 %
Ac x 3	18.2 %	25.3 %	24.6 %	24.7 %
Ac x 4	10.9 %	12.6 %	16.8 %	18.0 %
Ac x 5	1.2 %	2.4 %	1.6 %	1.8 %

Table 3.11. Quantitation of hyperacetylated H2A C-terminal lysine residues for two bloodstream form and two procyclic form samples.

Nearly 50% of all C-terminal H2A's are acetylated one or more times. Approximately 25% of H2A peptides detected had three acetylated lysines, with up to 5 lysines observed to be acetylated simultaneously at low levels (~ 1.8%). This degree of hyperacetylation is interesting, taken that the H2A C-terminal tail is exposed at the nucleosome surface (Luger et al. 1997; Wolffe 1998). Hyperacetylation of the C-terminal tail will neutralize the positively charged lysine residues, shown to be important for DNA binding (Bao et al. 2004), contributing to disruption of DNA binding and likely resulting in less tightly packed, euchromatic chromatin. Recently, a bromodomain containing protein, TbBDF 2, was shown to bind hyperacetylated H2A.V N-terminal tails via the Kac binding pocket (Yang et al. 2017). It is unclear if this or a similar protein may bind to hyperacetylated H2A. Figure 3.6 shows the MS^2 spectra of the hyperacetylated H2A C-terminal peptide.

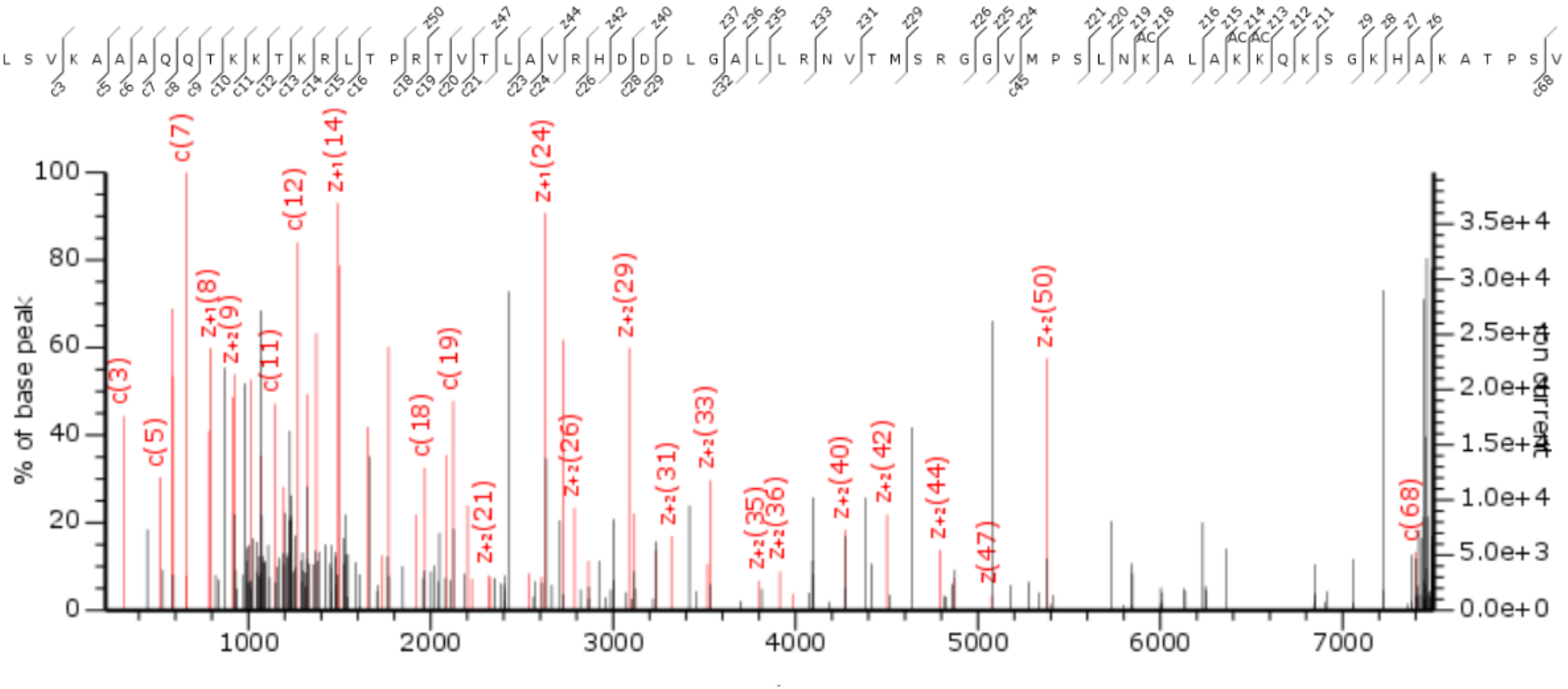


Figure 3.6: Annotated deconvoluted MS^2 spectra of the hyperacetylated C-terminal H2A peptide (residues 65 - 133) showing acetylated K115, K119, and K120.

Figure 3.7 provides a visual representation of the histone H2A N- and C-terminal peptide modification states.

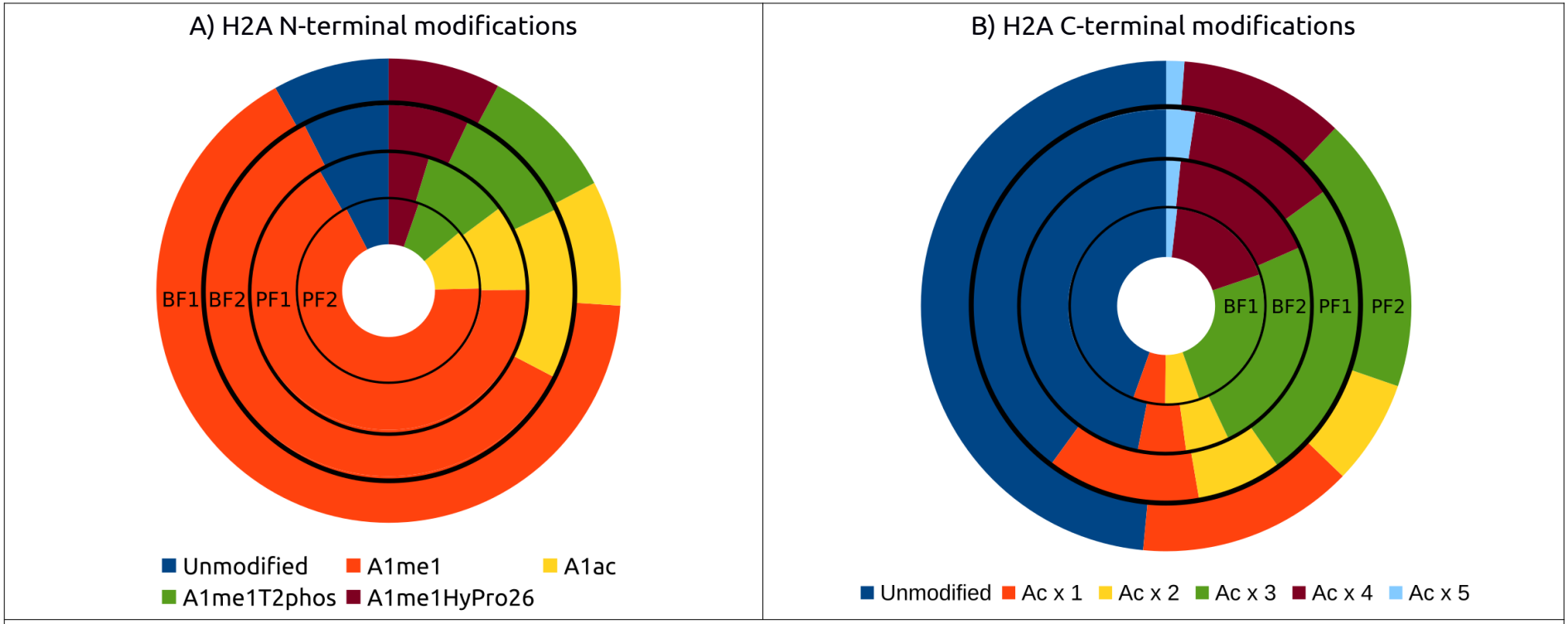


Figure 3.7: Visual representation of the quantification of H2A (**A**) N- and (**B**) C-terminal peptide modification.

The N-terminal peptides of histone H3 were sparsely modified, with only a few different PTMs detected. Although only a small number of PTMs were observed, most of the H3 peptides were found to be modified. About ~7% of BF H3 peptides and ~25% of PF H3 peptides did not carry any modifications, indicating that the N-terminal of H3 was modified to a greater extent in BF than PF cells. Histone H3 lysine 4 was observed to be acetylated, as previously seen (Wright, Siegel, and Cross 2010). Acetylation of H3K4 was previously found to be enriched at transcription start sites, with K4 acetylation mediated by the lysine acetyltransferase HAT4 (Siegel et al. 2008). It was also observed to be tri-methylated. Figure 3.8 shows the MS^2 spectra of tri-methylated H3K4.

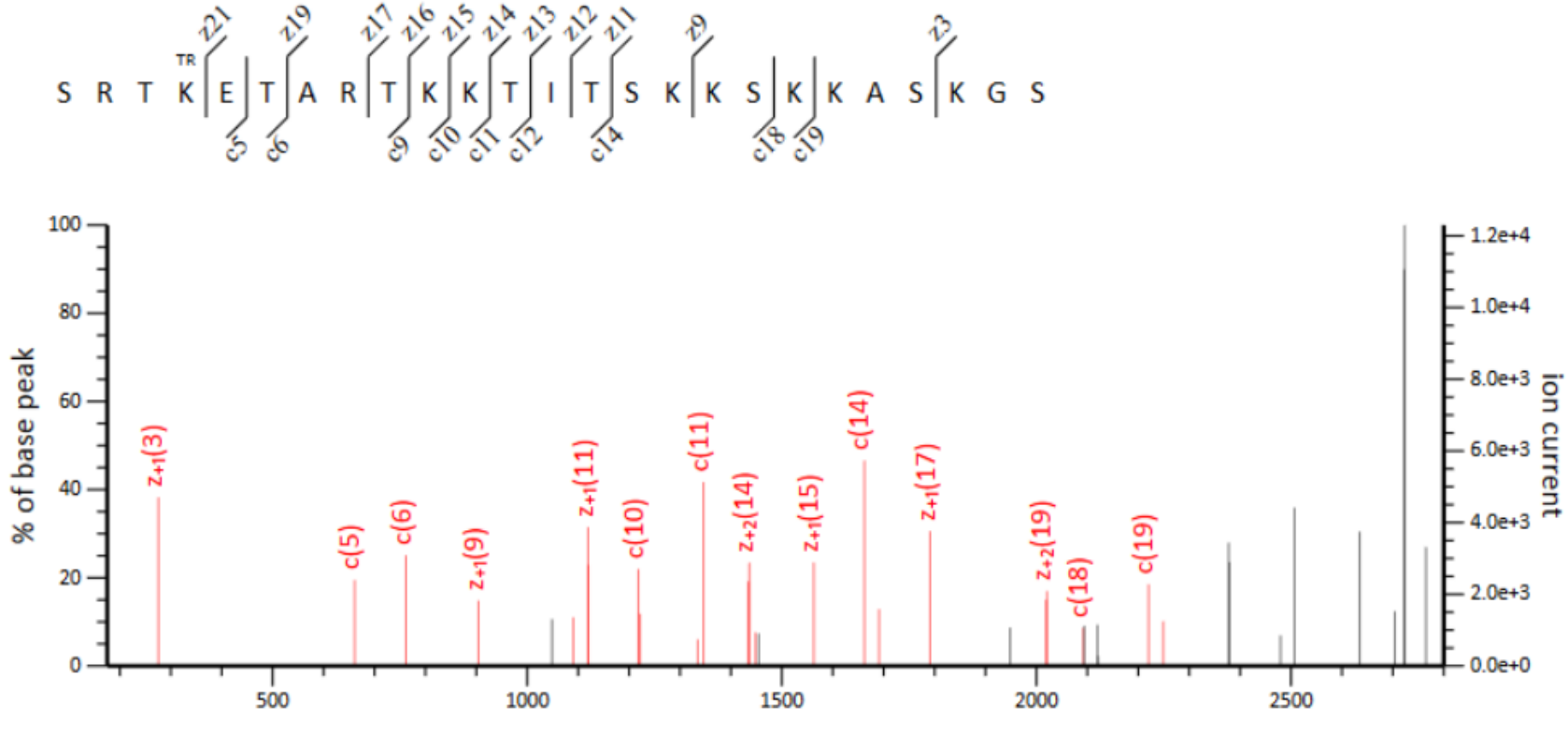


Figure 3.8. Deconvoluted MS^2 spectra of the H3K4 tri-methylated peptide (residues 1 - 26).

Intriguingly, tri-methylated H3K10 was found to be present in a cPTM pattern together with hydroxyproline 40 (discussed later on in section 3.4, see fig 3.11). Tri-methylated H3K10 has of yet not been observed in *T. brucei* or any related trypanosome species, with *T. cruzi* H3K10 being unmodified. In other eukaryotes, tri-methylated H3K9 is associated with transcriptional repression and heterochromatin formation (Lachner and Jenuwein 2002). However, it is unclear if *T. brucei* H3K10 is the functional equivalent residue as the sequence context is

notably different from H3K9 in other eukaryotes (Fig 3.9). The majority of H3 peptides were found to be hydroxylated at proline 40 (HyPro40), with ~80% of BFs and ~60% of PFs containing HyPro40. The difference in levels of HyPro40 in BF vs PF life cycles is an intriguing feature and may indicate that hydroxylation of H3 proline 40 is used to cycle-specifically mark genomic regions, thereby epigenetically regulating gene expression or transcription. Figure 3.10 shows a graphical representation of the histone H3 N-terminal modification states.

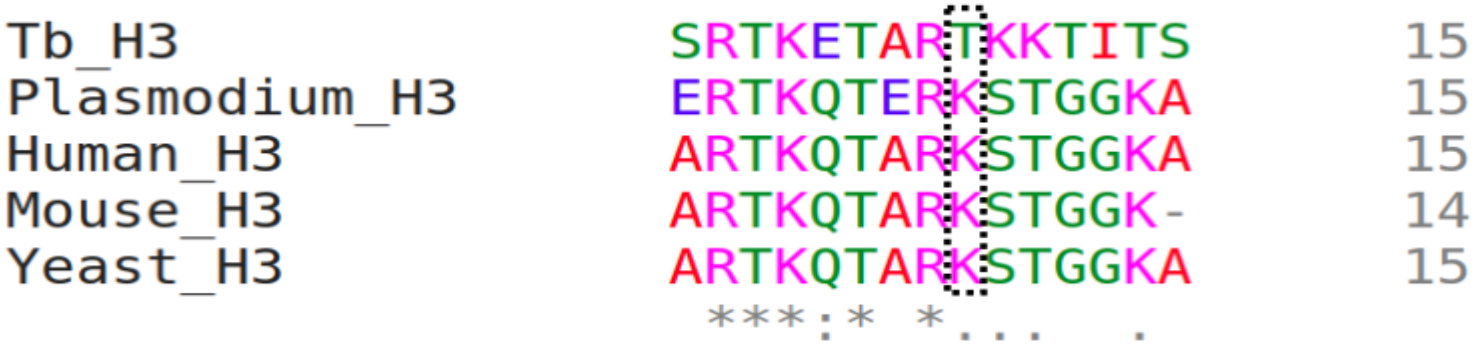


Figure 3.9. Sequence alignment of the first 15 amino acid residues in the N-terminal tail of histone H3 in *T. brucei* (Tb_H3), *Plasmodium falciparum* (Plasmodium_H3), *Homo sapiens* (Human_H3), *Mus musculus* (Mouse_H3), and *Saccharomyces cerevisiae* (Yeast_H3), showing the absence of H3K9 in *T. brucei.*

This binary H3K10me3HyPro40 PTM was detected at low levels, with only about ~1% of H3's displaying this mark in both life cycles (Table 3.12). Doubly oxidised H3 peptides were also observed (discussed in section 3.4). Figure 3.11 and 3.12 shows the MS^2 spectra of H3K10me3HyPro40 and H3HyPro40 peptides, respectively. Modification status of the N-terminal H3 peptides was quantitated as these PTMs were of interest to the current study (Tables 3.12).

Modification	**BF1**	**BF2**	**PF1**	**PF1**
Unmodified	6.8	6.6	27.0	22.3
HyPro40	85.0	82.8	58.4	61.8
$K_{10me3}HyPro_{40}$	0.9	1.1	0.7	0.8
$W_{38ox}HyPro_{40}$	7.3	9.5	13.8	15.0

Table 3.12. Quantitation of histone H3 N-terminal peptide modifications in two bloodstream form and two procyclic form samples.

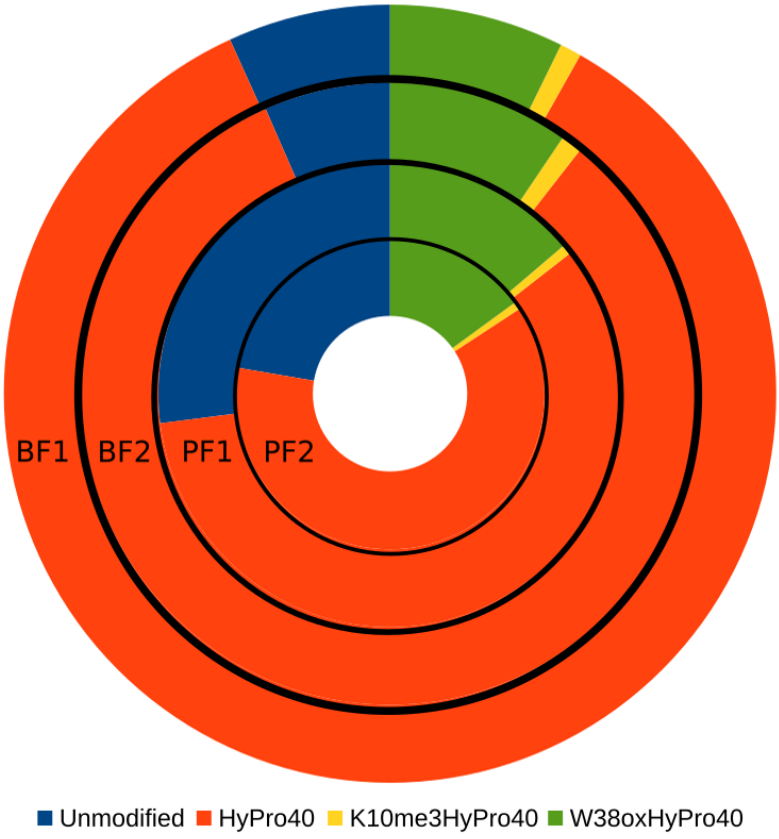


Figure 3.10: Visual representation of the quantitation of the H3 N-terminal peptide (resides 6 - 47) modifications, with unmodified in blue, HyPro40 in red, K10me3HyPro40 in yellow, and W38oxHyPro40 in green.

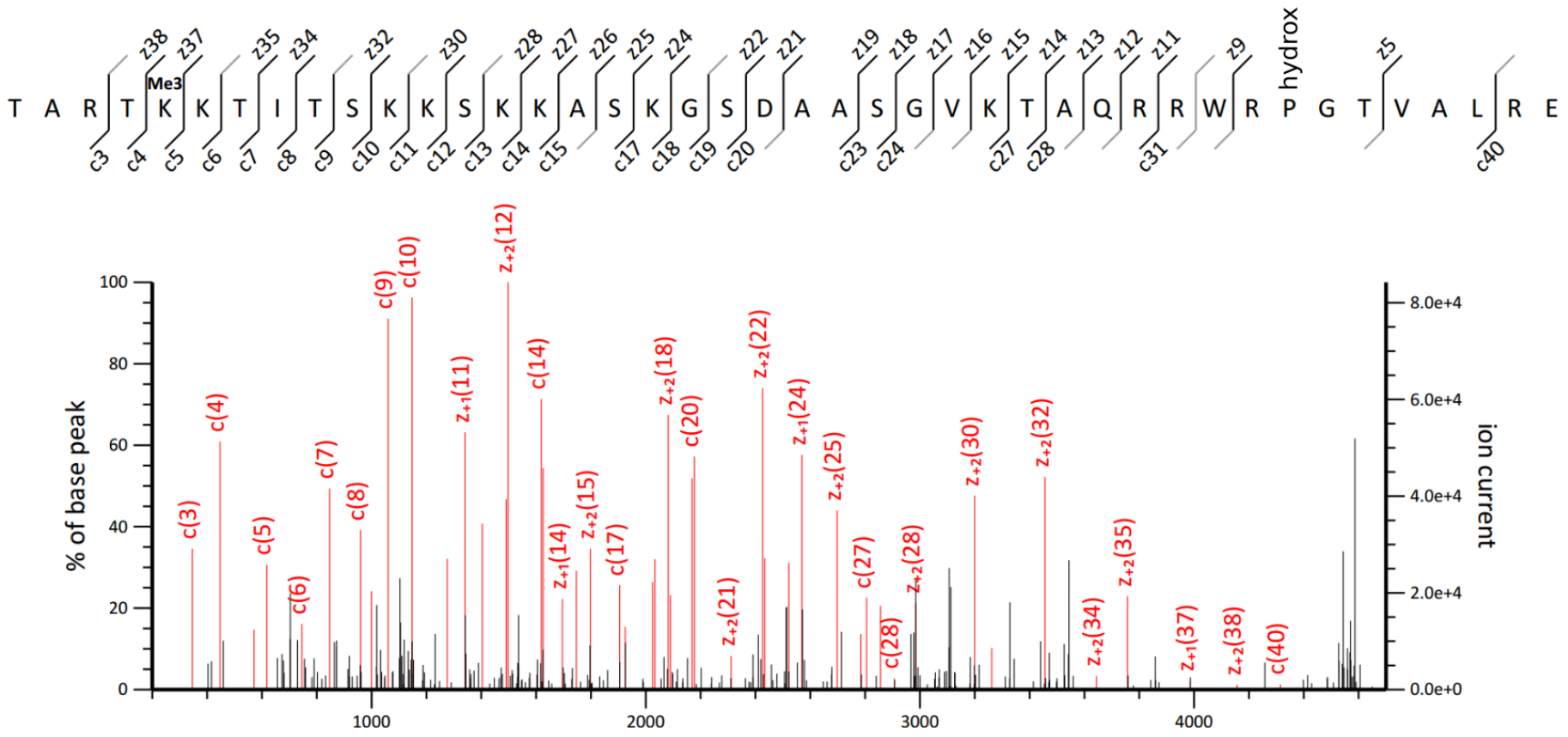


Figure 3.11. Annotated deconvoluted MS^2 spectrum of the N-terminal H3 peptide (residues 6 - 47) showing the H3K10me3HyPro40 binary PTM.

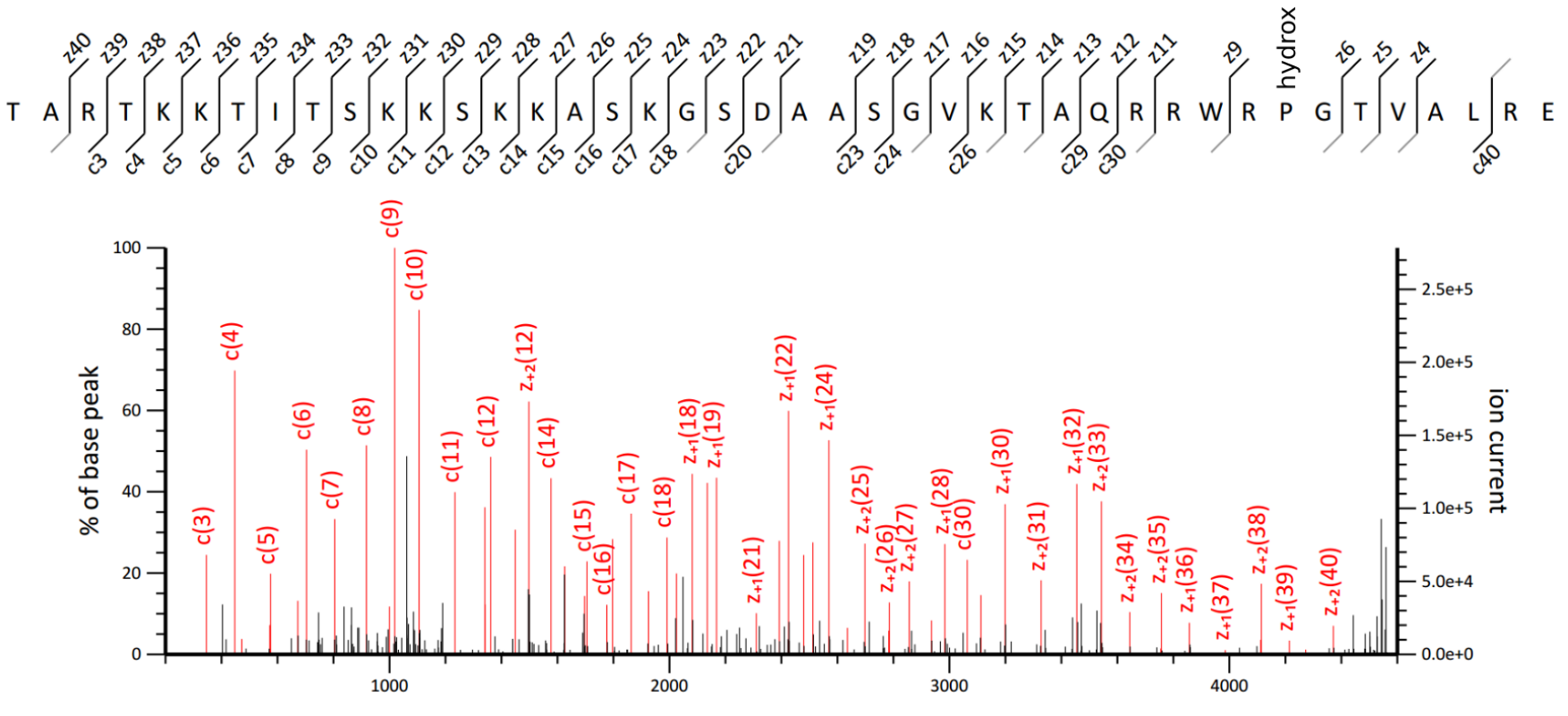


Figure 3.12. Annotated deconvoluted MS^2 spectrum of the H3 N-terminal peptide (residues 6 - 47) showing the spectra of the hydroxylated proline 40 peptide.

3.4) Hydroxyproline

An interesting feature detected in the MS spectra was the presence of a 16 *m/z* shift, corresponding to the mass shift of an oxidation or hydroxylation modification (Fig. 3.13). The mass shifts localized to two proline residues; one on histone H3 proline 40, the other on histone H2A proline 26, possibly present as 4-hydroxy-L-proline, a non-proteinogenic L-alpha amino acid.

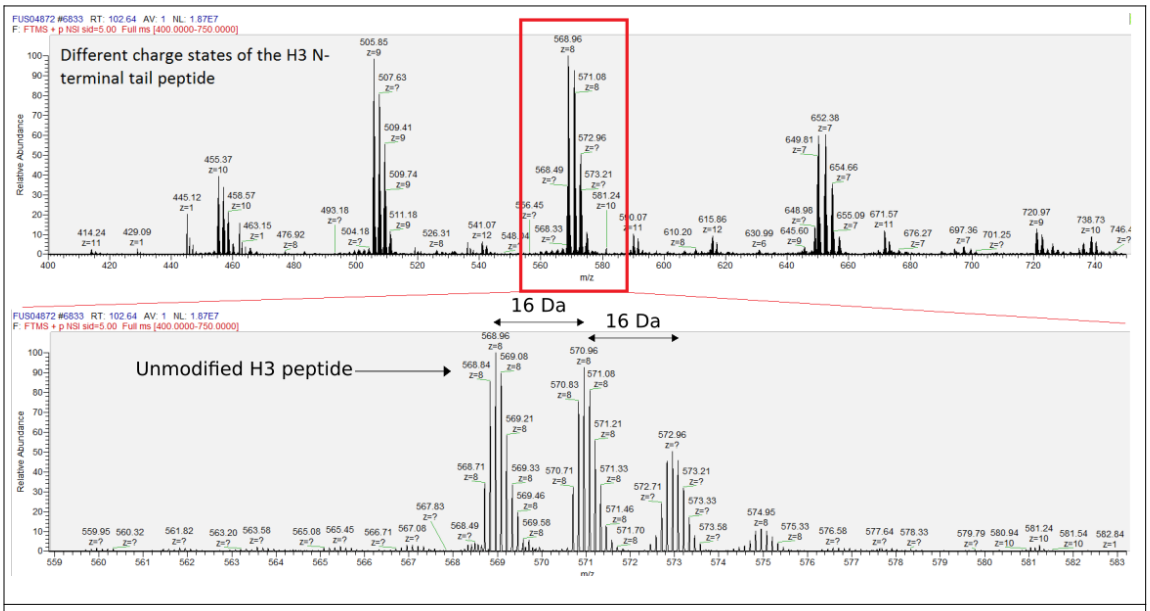


Figure 3.13. MS[1] spectra showing different charge states of the N-terminal histone H3 peptide (top panel), and the observed 16 Da mass shifts observed indicating the presence of oxidation and/or hydroxylation modifications (bottom panel).

H3HyPro40 was observed to exist as an individual PTM, and marks over 80% of H3 in BF and about 60% in PF samples (Fig. 3.12 and Table 3.12). H3HyPro40 was also observed as part of a binary modification (H3 K10me3HyPro40), marking ~1% in both BF and PF samples. Interestingly, H3K10me3 was observed to co-occur exclusively with H3HyPro40 (Fig 3.11). Doubly oxidised H3 peptides were also observed (W38oxHyPro40, Table 3.12). However, modification of tryptophan residues in all likelihood represented a sample artefact, as fortuitous oxidation of tryptophan has been observed to occur during sample handling or preparation (Perdivara et al. 2010; Verrastro et al. 2015).

Histone H2A Pro26 was also found to be hydroxylated, existing as a binary mark that co-occurred with A1me1, at ~7% in BF and ~5% in PF samples (Table 3.2). No other proline residues were found to be hydroxylated in any H2A and H3 peptides, with histone H3 proline 63 being unmodified.

To verify the chemical integrity of the samples used for MS analysis, and to address the question of whether H3HyPro40 and H2AHyPro26 residues were artefactual PTMs caused by fortuitous oxidation of proline residues, we repeated the histone extraction (with CTH spiked into *T. brucei* samples at the beginning or end of the protocol) and MS analysis (as described in section 2.7). MS analysis of CTH spiked samples once again revealed the presence of *T. brucei* H3HyPro40 and H2AHyPro26 residues,

yet no proline residues in the bovine histones were found to be hydroxylated. Notably, bovine H2APro26, the positional equivalent of *T. brucei* H2APro26 (Fig. 3.14), was unmodified, indicating that the hydroxyproline residues detected in the *T. brucei* histones were not the result of fortuitous oxidation during sample preparation.

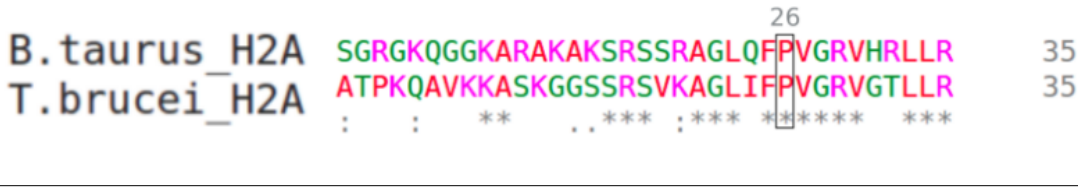


Figure 3.14: Clustal Omega alignment output of the first 35 amino acids of the N-terminal tail of *Bos taurus* and *T. brucei* histone H2A proteins.

To assess the theoretical possibility of proline hydroxylation on the above-mentioned residues, histone sequences were analyzed using RF-Hydroxysite, a bioinformatics tool that uses the primary amino acid sequence as well as physicochemical and structural data to predict proline hydroxylation (Ismail, Newman, and KC 2016). RF-Hydroxysite predicted two potential sites of proline hydroxylation on *T. brucei* histones, one on H2A proline 26 and one on H3 proline 40 (Fig. 3.15). These results strongly suggest that the hydroxyprolines observed at H2AHyPro26 and H3HyPro40 were *bona fide* post-translational modifications, and not a result of sample preparation.

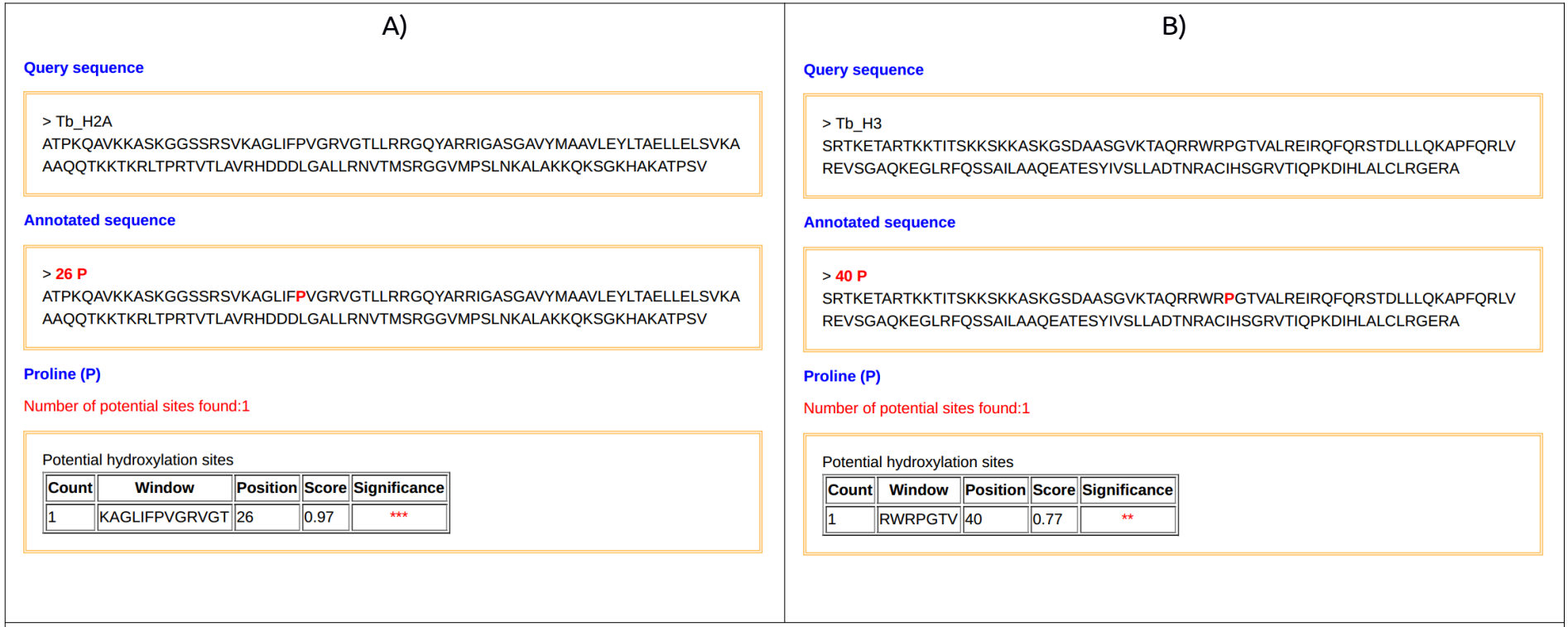


Figure 3.15. Proline hydroxylation sites as predicted by RF-Hydroxysite (**A**) Theoretical prediction of proline hydroxylation sites on *T. brucei* H2A pro 26, and (**B**) H3 pro 40. Potential HyPro sites indicated in red in the primary amino acid sequence. Significance: *** p<0.001, ** p<0.01.

3.5) Comparison of *T. brucei* cPTMs to other eukaryotes

As previously observed, H2A A1 was found to be N-terminally mono-methylated, a trypanosome specific modification also present in *T. cruzi* (Janzen et al. 2006; Picchi et al. 2017). However, acetylation of the A1 residue was not detected in *T. cruzi* and might be specific to *T. brucei*. Acetylation of H2A K4, previously observed in BF cells only, was detected in both BF and PF *T. brucei,* while mono-methylation of K4 residues appear to be conserved across the Trypanosome genus, although it was observed to be life-cycle specific in *T. cruzi* where it was only detected in the insect infective epimastigote stage (Picchi et al. 2017).

Hyperacetylation of the trypanosomal H2A C-terminal tail presents an intriguing peculiarity. Approximately 25% of all *T. brucei* H2A peptides displayed three acetylation marks, with up to five acetylated lysine residues detected simultaneously (Table 3.11). Hyperacetylation of the H2A C-terminal tail appears to be a conserved feature in trypanosomes, as similar hyperacetylation patterns were observed in *T. cruzi*. *T. brucei* H2A K115, K119, K120, K122, and K128 residues align with K116, K120, K121, K123, and K129 in *T. cruzi*, respectively (see Supplementary document R2). Comparable cPTM patterns were observed in these trypanosomes, with *T. brucei* K115acK119acK120ac and K119acK120ac cPTMs corresponding to K116acK120acK121ac and K120acK121ac cPTMs seen in *T. cruzi* (De Jesus et al. 2016; Picchi et al. 2017). These observations suggest that hyperacetylation of H2A C-terminal

lysine residues in specific patterns were established before trypanosomal speciation, and differential acetylation of these lysine residues may be used to serve distinct functions.

The histone H2A.V variant displayed only two PTMs that correlated to observed modifications in other trypanosomes. Acetylation of *T. brucei* H2A.V K55 and K59 correspond to K54ac and K58ac seen in *T. cruzi*, respectively (Picchi et al. 2017). Although only a few PTMs were confidently mapped to histone H2B, the two PTMs that were identified (mono-methylated K4 and K9) were novel to *T. brucei*, and have not been detected in a related trypanosome (Table 3.5).

In contrast to the H2A.V variant, H2B.V displayed complex modification patterns which localized to the N-terminal tail, similar to that seen on H2B.V in *T. cruzi,* with K4, K7, K15, K17, and K19 being acetylated or methylated to various degrees, although K19ac was not previously observed in *T. brucei* (Moretti et al. 2018). These modified H2B.V lysine residues are positionally equivalent in *T. brucei* and *T. cruzi,* and display strikingly similar cPTM patterns. Identical patterns of H2B.V N-terminal lysine hyperacetylation were observed in both trypanosomes, where the K7acK15acK17acK19ac cPTM observed in both *T. brucei* and *T. cruzi* (De Jesus et al. 2016; Picchi et al. 2017).

Although the trypanosomal H3 N-terminal tail diverges sharply from that seen in other eukaryotes (Fig. 1.4), lysine 4 is present in both *T. brucei* and *T. cruzi* H3 core histones, as well as higher eukaryotes like human, mouse and yeast H3. In its tri-methylated state, H3K4 is usually located at gene promoters and is associated with transcriptional activation in most eukaryotes (Sidoli, Cheng, and Jensen 2012; Garcia et al. 2007). In *T. brucei*, H3K4me3 was found to be enriched at putative pol II transcription start sites (Wright, Siegel, and Cross 2010). In this study, acetylated H3K4 was observed in *T. brucei* for the first time, although its function remains unclear. In *S. cerevisiae*, acetylated H3K4 was found to be enriched at promoters of actively transcribed genes and were located just upstream of H3K4me3 containing chromatin (Guillemette et al. 2011). No modification has been observed for the H3K4 residue in *T. cruzi* (De Jesus et al. 2016; Picchi et al. 2017).

Interestingly, H3K10 was found to be tri-methylated, a modification not previously observed on this residue in trypanosomes. In other eukaryotes, H3K9me3 is associated with transcriptional repression and heterochromatin formation (Lachner and Jenuwein 2002). However, it unclear if *T. brucei* H3K10 is the functional equivalent of H3K9 in other eukaryotes as the amino acid sequence context is markedly different (see Fig. 3.9). The function of H3K10me3 in *T. brucei* remains unclear.

Quantitation revealed that only a small percentage of H3 peptides were tri-methylated at K10 (~1%) which may correspond to the small proportion of the *T. brucei* genome being transcriptionally repressed. No modification of H3K11 residues were observed. Mono-methylation of H3K23 has been observed in *T. cruzi* where it was found to be enriched in non-proliferative life stages (De Jesus et al. 2016). In humans, mono-methylated H3K27, a possible homologue of *T. brucei* H3K23, localized to active promoters and was found to be elevated throughout transcriptionally active genes compared to the surrounding chromatin environment (Ahn et al. 2016). H3 lysine 76 was observed to be mono-, di-, or tri-methylated, as previously reported (Janzen et al. 2006; Mandava et al. 2007). Mono- and di-methylation of K76 residues are catalyzed by DOT1, a methyltransferase essential for cell cycle regulation, and occur only in late G2 and M cell-cycle phases in *T. brucei* (Janzen et al. 2006; Gassen et al. 2012). Tri-methylated H3K76 was found to be enriched in nucleosomes containing H3K4me3, H2A.V and H2B.V - chromatin marks associated with active transcription (Kouzarides 2007; Mandava, Janzen, and Cross 2008). Acetylation of H3K76 residues, previously reported in *T. cruzi*, was observed in *T. brucei* for the first time in this study.

Acetylation of histone H4K4 and K10 has been detected in *T. brucei*, *T. cruzi*, and *Leishmania donovani* (De Jesus et al. 2016; Picchi et al. 2017; Jha et al. 2017). In *T. brucei*, acetylation of H4K4 and K10 is

mediated by the lysine acetyltransferases HAT3 and HAT2, respectively (Siegel et al. 2008). Acetylation of K4 occurs after chromatin assembly and was found to be cell-cycle regulated. In *L. donovani*, acetylation of H4K4 residues is mediated by the histone acetyltransferase HAT2 and affects chromatin accessibility and compaction (Jha et al. 2017). *T. brucei* H4K10ac was found to localize at sites of putative pol II transcription initiation (Siegel et al. 2009). H4K14 was also observed to be acetylated, as previously seen (Mandava et al. 2007). Although the function of acetylated H4K14 in *T. brucei* remains unclear, H4K14ac was required for proper assembly and remodeling of chromatin during transcription and replication reactions in *T. cruzi* (Ramos et al. 2015). Immunofluorescence studies in *T. cruzi* revealed that acetylated H4 K10 and K14 residues were present at regions of relaxed chromatin, known sites of active transcription (Sheila Cristina Nardelli et al. 2009).

Amino acid modification by hydroxylation presents a subtle mass shift that adds a mere 16 atomic mass units to the modified residue (Gorres and Raines 2010). In mammalian tissues, proline hydroxylation plays a vital role in the conformational stability of the collagen triple helix where it is present as 4-hydroxy-L-proline (Jenkins et al. 2003), and is catalysed by prolyl-4-hydroxylase (P4H) enzyme (Myllyharju 2003). In *Dictyostelium*, a P4H enzyme has been observed to be crucial for hydroxylation of proline 143 of Skp1, a

small protein found in protein complexes in the cytoplasm and nucleus of eukaryotes (van der Wel et al. 2002). However, no genomic evidence has been found for the presence of any prolyl-4-hydroxylases in trypanosomes (West and Blader 2015).

It is unclear if proline hydroxylation will induce a conformational change between trans and cis configurations. In *S. cerevisiae*, cross-talk between H3P38 and H3K36 residues has been observed, where a conformational change induced by P38 isomerization resulted in structural changes in the H3 N-terminal tail required for H3K36 methylation by Set2, and subsequent transcriptional activation (Nelson, Santos-Rosa, and Kouzarides 2006).

In *T. brucei*, hydroxylation was detected on H2APro26 and H3Pro40, both of which formed part of a binary PTM, and was observed at similar levels between life-cycles. Investigation of fortuitous oxidation of *T. brucei* histone samples was done by addition of CTH at different stages of the histone extraction procedure, and subsequent analysis by LC-MS/MS, verified the chemical integrity of the *T. brucei* samples. *T. brucei* H2AHyPro26 and H3HyPro40 peptides were detected in these CTH spiked samples, with no mass shifts observed on proline residues of bovine histone peptides. Although hydroxylation modifications were observed on tyrosine residues in *T. cruzi*, no proline modifications were observed in this related trypanosome (Picchi et al. 2017). It is unclear what the

function of these (c)PTMs are in *T. brucei*, although their presence hints at the possible presence of a yet unidentified P4H in this trypanosome.

Taken together, this data points towards the existence of a robust epigenetic "histone code" employed by *T. brucei* to regulate genomic processes, and expands our understanding of the histone epigenetic repertoire used by *T. brucei*. The existence of homologous (c)PTMS observed in kinetoplasts and higher eukaryotes suggest that a (rudimentary) "histone code" existed in the eukaryotic progenitor before kinetoplasts diverged from the main eukaryotic lineage 1.5 billion years ago.

4.0 Investigation of genomic distribution of hyperacetylated histone H2A via MNase-ChIP-seq.

Over the past few decades, investigation into trypanosome epigenetics has revealed a myriad of epigenetic signals and effector proteins, some of which have been mapped to specific, and often crucial functions. In trypanosomes, histone epigenetics have been increasingly scrutinized as it has become apparent that these organisms use epigenetic mechanisms to delineate its genome, and control DNA function. In *T. brucei*, several PTMs have been identified on core histones (Mandava et al. 2007; Siegel et al. 2009). On histone H2A, the C-terminal was found to contain multiple acetylated lysines (K115, 119, 120, 122, 125, 128), of which some (K120, 122, and 128) correspond to conserved lysine residues with defined epigenetic functions. At regions of pol II transcription initiation, tri-methylated H3K4 and acetylated H4K10 were found to co-localize with histone variants H2A.V and H2B.V and the chromatin effector protein BDF3.

Looking at the similarity of histone PTMs in trypanosomes and the vast number of (c)PTMs identified in *T. cruzi* (Picchi et al. 2017), it is probable that *T. brucei* may employ an equally intricate assemblage of histone combinatorial PTM patterns to demarcate and delineate its genome. Investigation of the distribution of cPTMs by MNase-ChIP-seq will undoubtedly unveil the epigenetic function of selected cPTMS and further the understanding of the *T. brucei* epigenetic repertoire.

4.1) Antibody cross-reactivity analysis

ChIP quality antibodies were raised in New Zealand rabbits by peptide inoculation as discussed under section 2.8. Test bleeds were taken after the 3rd immunization, affinity purified and antibody specificity determined via indirect ELISA (GenScript Incorporated, Piscataway, USA).

α-H2AT2pho antibody:

Mono-methylation of H2A Ala1 has been observed to independently exist on the N-terminal of histone H2A previously (Mandava et al. 2007) as well as in this study. The A1me1 modification was included in the unmodified peptide to prevent pull-down of both A1me1 and A1me1T2pho containing chromatin during ChIP and prevent artefactual enrichment of chromatin only bearing the A1me1 mark.

- Modified peptide: $A_{\mathbf{1me1}}T_{\mathbf{2pho}}$PKQAVKKASK
- Unmodified peptide: $A_{\mathbf{1me1}}T_2$PKQAVKKASK

	Titer	[Ab] ng/ml	Modified	Unmodified
Negative Ctrl (NC)	1:1,000	N/A	0.121	0.228
1	1:1,000	1,000	3.639	0.249
2	1:2,000	500	3.566	0.164
3	1:4,000	250	3.528	0.120
4	1:8,000	125	3.374	0.107
5	1:16,000	62.5	3.193	0.103
6	1:32,000	31.25	2.889	0.099
7	1:64,000	15.62	2.275	0.099
8	1:128,000	7.81	1.537	0.095
9	1:256,000	3.90	0.937	0.094
10	1:512,000	1.95	0.532	0.092
11	Blank	Blank	0.069	0.097
12	Blank	Blank	0.069	0.097
Titer			>1:512,000	~1:1000

Table 4.1: Indirect ELISA with α-H2AT2pho antibody using modified and unmodified peptides.

Following QC and indirect ELISA assay, antibody cross-reactivity was examined via Western Dot Blot analyses against modified and unmodified peptides to manually examine antibody cross-reactivity using the Pierce™ Fast Western Blot Kit (Thermo Scientific). Five micrograms of each peptide was blotted on a nitrocellulose membrane, dried for 10 minutes to fix, inoculated with the antibody (1:10 000 dilution), and visualised on a MyECL. No cross-reactivity was detected (Fig. 4.1).

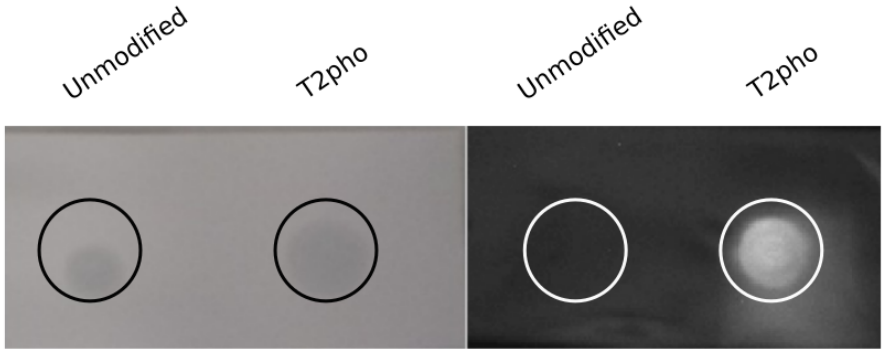


Figure 4.1. Dot Blot analysis to assess cross-reactivity of α-H2AT2pho antibody: Five micrograms of the unmodified and modified peptides were applied to a nitrocellulose membrane and probed with α-H2A T2pho antibody. No cross-reactivity was observed. Blots are indicated on the left, MyECL exposure on the right.

α-H2AK115ac antibody:

As the hyperacetylated C-terminal tail of histone H2A may exist in a few combinatorial PTM patterns with up to five acetylated lysines detected during MS analysis, it was essential that the antibodies raised against H2AK115ac and H2AK125ac did not cross-react, or reacted with acetylated H2AK120acK122ac. Indirect ELISAs as well as western dot blot analysis confirmed that the antibodies raised selectively bound to their respective marks, and did not cross-react with the unmodified peptide (see below).

•H2AK115ac modified peptide:LN**K_{115ac}**ALAK**K_{120ac}**Q**K_{122ac}**SGK_{125}HAC
•H2AK125ac modified peptide:LNK_{115}ALAK**K_{120ac}**Q**K_{122ac}**SG**K_{125ac}**HAC
•Unmodified peptide: LNK_{115}ALAKK_{120}QK_{122}SGK_{125}HAC

	Titer	[Ab] ng/ml	K115ac	K125ac	Unmod
NC	1:1,000	N/A	0.056	0.089	0.106
1	1:1,000	1,000	3.184	0.200	0.225
2	1:2,000	500	2.938	0.104	0.131
3	1:4,000	250	2.669	0.104	0.131
4	1:8,000	125	2.274	0.056	0.076
5	1:16,000	62.5	1.736	0.047	0.066
6	1:32,000	31.25	1.159	0.048	0.061
7	1:64,000	15.62	0.691	0.048	0.061
8	1:128,000	7.81	0.401	0.044	0.061
9	1:256,000	3.90	0.188	0.044	0.063
10	1:512,000	1.95	0.150	0.042	0.068
11	Blank	Blank	0.059	0.046	0.065
12	Blank	Blank	0.059	0.046	0.065
Titer			>1:512,000	1:2,000	~1:1000

Table 4.2. Indirect ELISA assay against α-H2AK115ac antibody using modified (K115ac and K125ac) and unmodified peptides.

α-H2AK125ac antibody:

- H2AK115ac Modified peptide: $LNK_{115ac}ALAKK_{120ac}QK_{122ac}SGKHA$
- H2AK125ac Modified peptide: $LNKALAKK_{120ac}QK_{122ac}SGK_{125ac}HA$
- Unmodified peptide: $SLNK_{115}ALAKK_{120}QK_{122}SGK_{125}HAC$

	Titer	[Ab] ng/ml	K125ac	K115ac	Unmod
NC	1:1,000	N/A	0.062	0.081	0.099
1	1:1,000	1,000	3.448	0.337	0.127
2	1:2,000	500	3.413	0.174	0.104
3	1:4,000	250	3.262	0.114	0.099
4	1:8,000	125	2.981	0.079	0.092
5	1:16,000	62.5	2.761	0.063	0.088
6	1:32,000	31.25	2.324	0.056	0.087
7	1:64,000	15.62	1.773	0.053	0.084
8	1:128,000	7.81	1.105	0.052	0.080
9	1:256,000	3.90	0.616	0.051	0.079
10	1:512,000	1.95	0.347	0.05	0.078
11	Blank	Blank	0.053	0.053	0.085
12	Blank	Blank	0.053	0.053	0.085
Titer			>1:512,000	1:4,000	~1:2000

Table 4.3: Indirect ELISA assay against α-H2AK125ac antibody using modified (K115ac and K125ac) and unmodified peptides.

Antibody cross-reactivity was examined via Western Dot Blot analyses against the unmodified and modified peptides with α-K115ac or α-K125ac antibodies, and visualised on a MyECL (Thermo Scientific). No cross-reactivity was detected (Fig. 4.2).

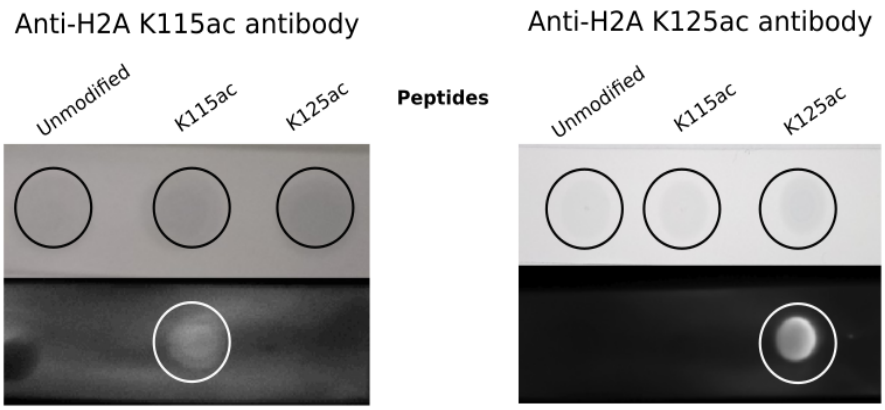


Figure 4.2. Dot Blot analysis to assess cross-reactivity of α-H2AK115ac and α-H2AK125ac antibodies: Five micrograms of the unmodified, K115ac, or K125ac peptides were applied to a nitrocellulose membrane and probed with either α-H2AK115ac (left-hand blot) or α-H2AK125ac (right-hand blot). Blots are indicated on the top, MyECL exposure(chemiluminescence) on the bottom.

α-H3K4me3 and α-H3K10me3 antibodies:

Antibodies were raised against H3K4me3 and H3K10me3 modifications in four and three different rabbits, respectively. Although both antibodies showed low levels of cross-reactivity with the unmodified H3 peptide, they exhibited high cross-reactivity for both of the modified peptides. Depletion of cross-reacting antibodies was attempted by several rounds of cross-adsorption. However, we were unable to obtain anti-H3 K4me3 and K10me3 specific antibodies as the antibodies were unable to differentiate between these marks. This might be as a result of the sequence context surrounding the modifications being remarkably similar, with both modifications containing an "RTK$_{me3}$" region (SRTK$_{me3}$E for H3K4me3, and ARTK$_{me3}$K for H3K10me3).

It is possible that the "RTK$_{me3}$" region is highly antigenic and may, therefore, induce production of analogous antibodies. These antibodies were not sequence-dependent and were likely to recognise any tri-methylated lysine. The similarity between the two marks caused high cross-reactivity between modified peptides and raised antibodies (Tables 4.4 and 4.5) and, therefore, these antibodies were not used to perform chromatin immunoprecipitation experiments.

<u>α-H3K4me3 antibody:</u>

- H3K4me3 peptide: SRT**K$_{4me3}$**ETARTK$_{10}$KTITSKKC (K4me3)
- H3K10me3 peptide: SRTK$_4$ETART**K$_{10me3}$**KTITSKKC (K10me3)
- Unmodified peptide: SRTK$_4$ETARTK$_{10}$KTITSKKC (Unm)

Rabbit no:	NC	1	2	3	4	5	6	7	8	9	10	11		
	1:1,000	1:1,000	1:2,000	1:4,000	1:8,000	1:16,000	1:32,000	1:64,000	1:128,000	1:256,000	1:512,000	Blank	Titer	Peptide
#6451	0.065	0.737	0.369	0.205	0.132	0.127	0.078	0.073	0.066	0.058	0.056	0.065	1:4,000	K4me3
	N/A	0.716	0.310	0.183	0.120	0.102	0.094	0.083	0.081	0.077	0.067	0.070	1:4,000	K10me3
	0.070	0.224	0.136	0.095	0.076	0.067	0.066	0.065	0.059	0.058	0.054	0.068	1:1,000	Unm
#6452	0.059	2.815	2.109	1.433	0.819	0.414	0.228	0.136	0.097	0.069	0.068	0.065	1:32,000	K4me3
	N/A	2.826	2.089	1.469	0.864	0.428	0.229	0.141	0.100	0.073	0.068	0.070	1:32,000	K10me3
	0.060	0.084	0.067	0.061	0.061	0.061	0.060	0.060	0.059	0.059	0.057	0.068	<1:1,000	Unm
#6453	0.101	2.973	2.415	1.801	1.078	0.558	0.297	0.180	0.114	0.082	0.072	0.065	1:64,000	K4me3
	N/A	3.028	2.330	1.672	1.081	0.590	0.319	0.179	0.117	0.092	0.075	0.070	1:64,000	K10me3
	0.103	0.274	0.142	0.096	0.079	0.070	0.063	0.062	0.061	0.061	0.057	0.068	1:1,000	Unm
#6454	0.082	2.795	2.323	2.128	1.612	1.024	0.565	0.324	0.189	0.108	0.087	0.065	1:128,000	K4me3
	N/A	2.960	2.282	1.619	0.981	0.550	0.301	0.177	0.119	0.093	0.077	0.070	1:64,000	K10me3
	0.086	0.117	0.089	0.079	0.074	0.069	0.068	0.066	0.065	0.064	0.063	0.068	<1:1,000	Unm

Table 4.4: Indirect ELISA against modified and unmodified peptides with the α-H3K4me3 antibody.

α-H3K10me3 antibody:

- H3K4me3 peptide: SRT**K$_{4me3}$**ETARTK$_{10}$KTITSKKC (K4me3)
- H3K10me3 peptide: SRTK$_4$ETART**K$_{10me3}$**KTITSKKC (K10me3)
- Unmodified peptide: SRTK$_4$ETARTK$_{10}$KTITSKKC (Unm)

Rabbit no:	NC	1	2	3	4	5	6	7	8	9	10	11		
	1:1,000	1:1,000	1:2,000	1:4,000	1:8,000	1:16,000	1:32,000	1:64,000	1:128,000	1:256,000	1:512,000	Blank	Titer	Peptide
#6760	0.138	2.416	1.498	0.861	0.487	0.259	0.167	0.114	0.111	0.087	0.079	0.070	1:32,000	K4me3
	0.158	1.634	0.936	0.531	0.329	0.191	0.158	0.144	0.114	0.113	0.097	0.089	1:16,000	K10me3
	0.151	1.281	0.624	0.405	0.235	0.197	0.173	0.150	0.128	0.109	0.102	0.101	1:8,000	Unm
#6761	0.074	2.542	1.942	1.295	0.689	0.368	0.262	0.160	0.111	0.087	0.076	0.070	1:64,000	K4me3
	0.089	2.146	1.439	0.915	0.522	0.371	0.211	0.155	0.105	0.105	0.087	0.089	1:32,000	K10me3
	0.199	0.134	0.134	0.104	0.095	0.086	0.082	0.081	0.081	0.081	0.078	0.101	<1:1,000	Unm
#6762	0.077	2.939	2.698	2.279	1.600	0.889	0.533	0.275	0.174	0.117	0.091	0.070	1:128,000	K4me3
	0.110	2.840	2.469	1.967	1.326	0.796	0.452	0.290	0.190	0.119	0.118	0.089	1:128,000	K10me3
	0.133	0.293	0.188	0.174	0.150	0.110	0.104	0.089	0.086	0.086	0.068	0.101	1:1,000	Unm

Table 4.5 Indirect ELISA against modified and unmodified peptides with the α-H3K10me3 antibody.

α-H3 Hydroxyproline 40 antibody:

Raising H3 HyPro40 specific antibodies was unsuccessful as antibodies showed high cross-reactivity with the unmodified H3Pro40 peptide, and cross-adsorption did not yield a HyPro40 specific antibody. This may be due to structural similarities between these residues, with proline and 4-hydroxyproline only differing in one -OH group. Antibodies were unable to distinguish between modified and unmodified peptides (Table 4.6).

- H3HyPro40 peptide: CKTAQRRWR**HyPro$_{40}$**GTVA (A)
- H3Pro40 unmodified: CKTAQRRWRPro$_{40}$GTVA (B)

	Dilution	[Ab] ng/ml	Rabbit #6423		Rabbit #6424		Rabbit #6425		Rabbit #6426	
			A	B	A	B	A	B	A	B
NC	1:1,000	N/A	0.069	0.057	0.175	0.127	0.071	0.074	0.089	0.061
1	1:1,000	1,000	3.487	3.667	3.285	3.266	3.301	3.226	3.502	3.486
2	1:2,000	500	3.468	3.383	2.614	2.735	2.879	2.909	3.328	3.324
3	1:4,000	250	3.214	3.114	1.899	1.950	2.290	2.249	2.995	3.059
4	1:8,000	125	2.796	2.658	1.032	1.192	1.456	1.491	2.459	2.505
5	1:16,000	62.5	2.081	1.888	0.509	0.560	0.695	0.742	1.624	1.861
6	1:32,000	31.25	1.323	1.133	0.277	0.335	0.390	0.394	0.946	1.061
7	1:64,000	15.62	0.636	0.561	0.168	0.180	0.200	0.209	0.467	0.580
8	1:128,000	7.81	0.332	0.310	0.112	0.115	0.130	0.136	0.261	0.328
9	1:256,000	3.90	0.178	0.174	0.080	0.080	0.102	0.092	0.155	0.189
10	1:512,000	1.95	0.115	0.116	0.078	0.066	0.091	0.072	0.108	0.124
11	Blank	Blank	0.059	0.057	0.059	0.057	0.059	0.057	0.059	0.057
12	Blank	Blank	0.059	0.057	0.059	0.057	0.059	0.057	0.059	0.057
Titer			1:256000	1:256000	1:64000	1:64000	1:128000	1:128000	1:256000	1:256000

Table 4.6: Indirect ELISA against modified and unmodified peptides with α-H3HyPro40**.**

4.2) MNase-ChIP-seq and bioinformatic data processing

MNase is a nonspecific endo- exonuclease capable of digesting unbound dsDNA, leaving nucleosome-bound DNA fragments of ~147 bp intact (Cole, Howard, and Clark 2012). Normally, mono-nucleosomal DNA fragments are isolated by agarose gel purification from higher order nucleosomal DNA before sequencing. However, this approach was not used in this study as histone PTMs can alter the state of chromatin compaction, leading to over or under digestion of DNA. Selection of mono-nucleosomal size DNA fragments by gel purification may, therefore, introduce artefacts as the effect of the observed PTMs investigated in this study on *T. brucei* chromatin stability and compaction was unknown. Trypanosome cells were permeabilised using mild detergents, and the cytosolic content removed with consecutive washes. This approach results in trypanosome "ghosts"; a porous cellular membrane removed of most of the cytosolic content and organelles, but contains an intact cell nucleus. The nuclear envelope is then permeabilised with a mild detergent, chromatin digested with MNase to yield nucleosome core particles, and selected PTMs immunoprecipitated using custom made antibodies.

All bioinformatics data analyses were performed using the Galaxy platform (Afgan et al. 2018). *Trim_Galore!* was used to remove Illumina adapter sequences, filter DNA reads to select fragments of 115 - 125 nt in length, and remove low quality reads, whereafter *FastQC* was used to assess the read quality (Andrews 2010). Figure 4.3 shows the *FastQC* output for sample 3A (histone H3 PF input) as representative.

Quality control yielded an average fragment distribution length of 125 nt (Fig. 4.4). Paired-end sequence realignment using *Bowtie 2* (Langmead et al. 2009) resulted in an average alignment rate of ~70% for input, and ~63% for ChIP samples (Table 4.7). This alignment rate is in accordance with previous findings (Maree et al. 2017), as the intermediate- and mini-chromosome sequences as well as mitochondrial DNA constitutes approximately 30% of the total DNA complement (Vargas-Parada 2010), and does not form part of the current reference genome.

Sample no:name	**Description**	**Alignment rate (%)**
1: BF H3	Input	68
	ChIP	67
3: PF H3	Input	72
	ChIP	78
17: BF1 K125	Input	67
	ChIP	58
18: BF2 K125	Input	66
	ChIP	49
19: PF1 K125	Input	66
	ChIP	61
20: PF2 K125	Input	67
	ChIP	52
21: BF K115	Input	72
	ChIP	69
24: PF K115	Input	72
	ChIP	70

Table 4.7: Summary of paired-end sequence alignments by *Bowtie2*.

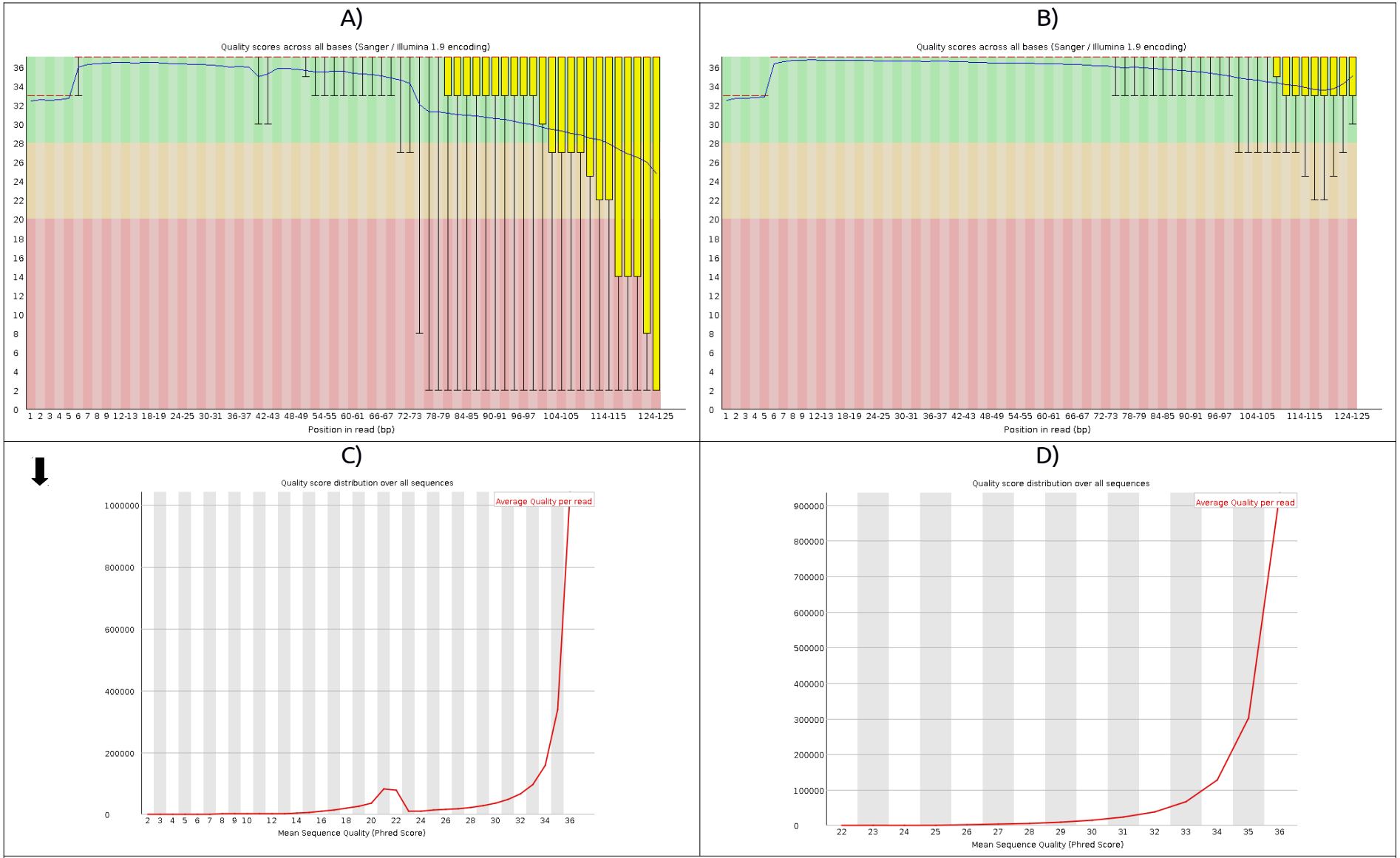


Figure 4.3. *FastQC* output showing results of read quality trimming by *Trim_Galore!*: (**A** and **B**) Per base and (**C** and **D**) per sequence quality scores before and after read trimming, respectively.

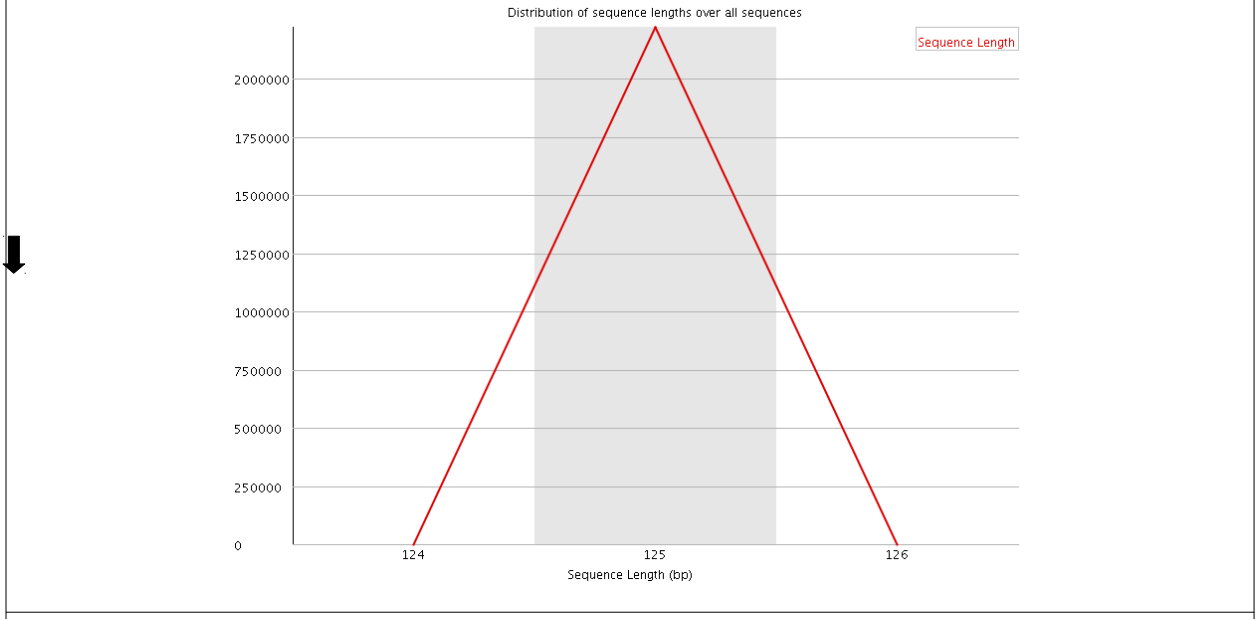


Figure 4.4. FastQC output showing the average sequence length distribution.

To identify genomic loci enriched for specific combinatorial PTM patterns, *callpeak*, a function of the *MACS2* package, was used to evaluate the ChIP-seq data, and identified peaks of enrichment visualized using IGV (Robinson et al. 2011; Thorvaldsdóttir, Robinson, and Mesirov 2013). Figure 4.5 shows an IGV snapshot of the *MACS2* called peaks for the 11 megabase-chromosomes (MBCs) of the *T. brucei* genome. Called peaks were named according to sample and peak number, i.e. S17_42 translates to sample 17, peak number 42.

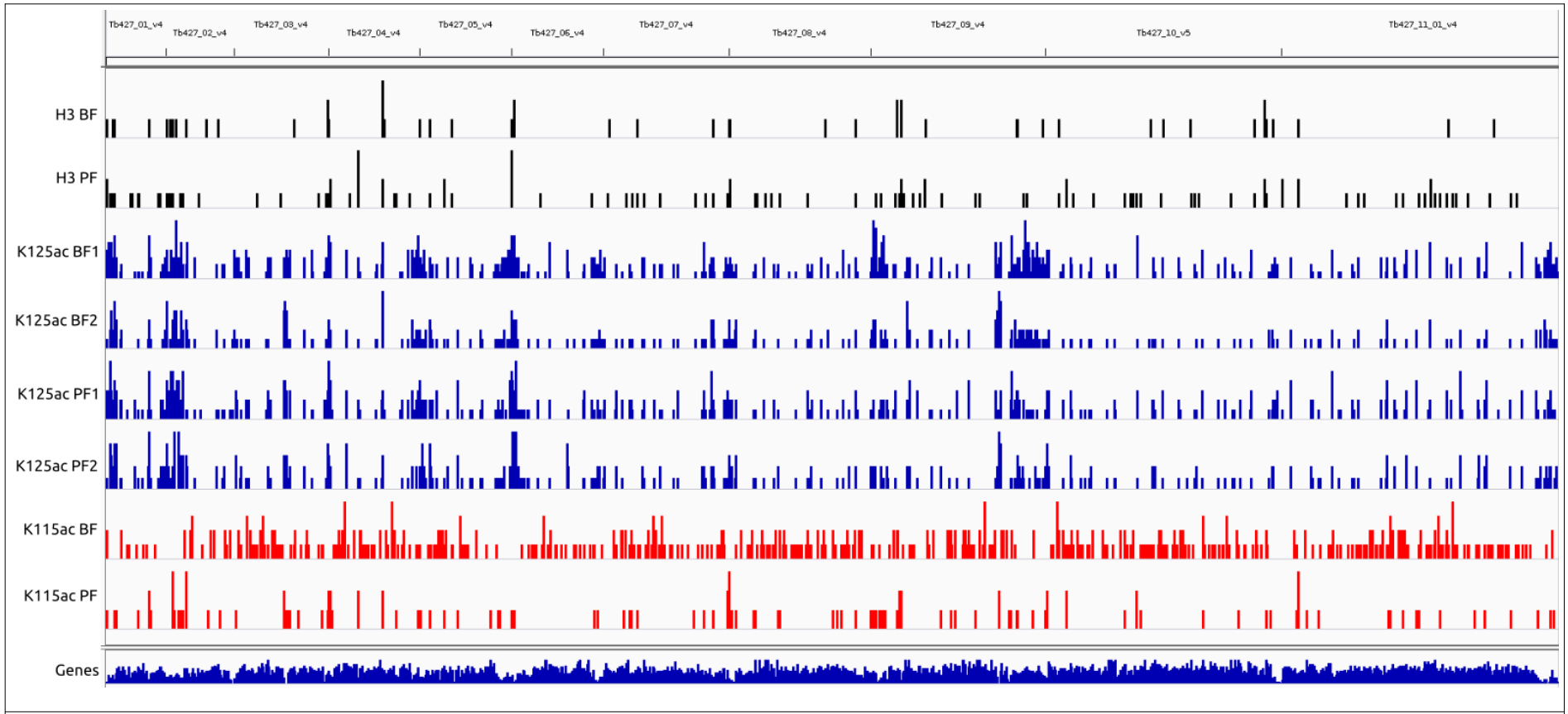


Figure 4.5: Genome-wide distribution of called peaks across the 11 *T. brucei* MBCs: Panels from top to bottom. (Top panel) the 11 MBCs of the *T. brucei* genome. (Black panels) BF and PF histone H3 peaks. (Blue panels); BF K125ac and PF K125ac peaks, replicates 1 and 2, respectively. (Red panels) BF and PF K115ac peaks. Individual bars indicate called peaks, with bar height indicating peak intensity. (Bottom panel) Annotated genes across the MBCs.

Visual inspection of enriched areas showed no striking difference between BF and PF life-forms. This is in agreement with previous findings where it was observed that nucleosomal density and placement were highly comparable between BF and PF life-forms (Maree et al. 2017), and the genome-wide distribution of acetylated H4K10 was also observed to be remarkably similar between BF and PF cells (Siegel et al. 2009).

Manual data inspection revealed a prominent co-localization of H2AK125ac peaks which showed enrichment at matching genomic loci in both BF and PF life-forms, and their respective biological replicates (Fig. 4.6, A). Perusal of the nucleosome positioning maps generated by *Bowtie2*, and subsequently *MACS2*, areas of peak enrichment became clearly visible and frequently coincided with peaks identified by MACS2 (Fig. 4.6, B)

To investigate the genomic distribution of the identified peaks, *intersect intervals* was used to find regions where combinatorial PTMs were enriched relative to specific genomic loci or regions. Because of the unique genomic arrangement of the *T. brucei* genome, the MBCs can be partitioned into several distinct/functional regions; e.g. genes transcribed by pol I/II/III, regions of pol II transcription initiation or termination, centromeric or sub-telomeric regions, etc. Table 2.4 (section 2.14, Genomic Region Intersect) summarizes the functional genomic partitioning used for peak intersection, based on previous work (Maree et al. 2017).

Figure 4.6. Colocalization of K125 peaks: (**A**) Representative region of chromosome 9 displaying colocalization of K125 peaks in BF and PF samples. (**B**) Nucleosomal positioning maps showing K125 enrichment in BF [BF 1 K125] and PF [PF 1 K125] samples as representatives. Nucleosome maps show juxtaposition of Input DNA in blue (above horizontal axis), and K125 ChIP DNA in red (below axis), with regions of K125 enrichment corresponding to *MACS2* called peaks.

4.3) Genomic distribution of PTMs

The trypanosome MBCs harbors ~7600 non-redundant protein-coding genes that are interrupted by various non-coding RNA genes that play an important role in different cellular processes. These RNAs are transcribed by different polymerases, and, when present in a pol II PTU, may present a steric hindrance to a transcribing pol II complex contributing to pol II transcriptional termination (Maree and Patterton 2014).

5S rRNAs

The pol III transcribed 5S rRNA gene cluster, organized as a single polycistronic unit on chromosome 8 (position 454171 - 460512), is located in a divergent pol II SSR. The 5S rRNA genes are identical, and fragment realignment by *Bowtie2* will not map individual genes, but distribute the signal across all genes. Looking at the ChIP nucleosomal profiles, all samples displayed a regularly spaced nucleosomal pattern across the entire 5S rRNA array, with the K125ac samples showing the highest enrichment (Fig 4.7). In the K125ac samples, regular nucleosomal patterns were seen on 5S rRNA array with greater levels of *non-K125ac* nucleosomes present *between* the 5S genes, and K125ac containing nucleosomes *on* the 5S genes themselves (Fig 4.7, Input vs ChIP). These regularly interspersed peaks have a periodicity of ~517 nt for input and ~542 nt in ChIP samples, with ChIP peaks being bordered on either side by NDRs (Fig. 4.8), indicative of the repetitive nature of the 5S rRNA array. This is in agreement with previous results showing three positioned nucleosomes on the 5S transcription unit, bordered by NDRs up- and downstream of these observed nucleosomes (Maree et al. 2017).

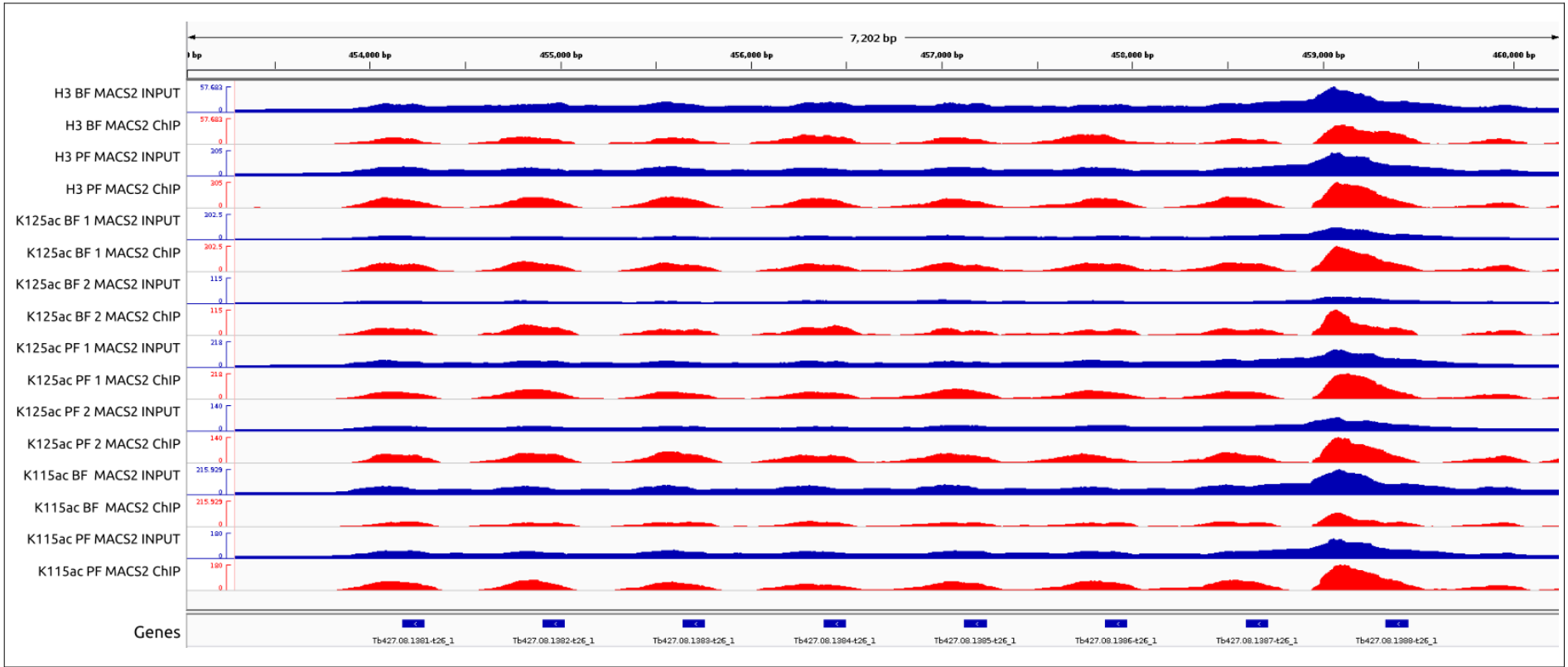


Figure 4.7: Nucleosome profiles for all INPUT (blue) and ChIP (red) samples across the 5S rRNA unit located towards the left-hand side of chromosome 8 at position 454171 to 460512. Regularly spaced nucleosomal patterns are observable across the entire 5S rRNA array.

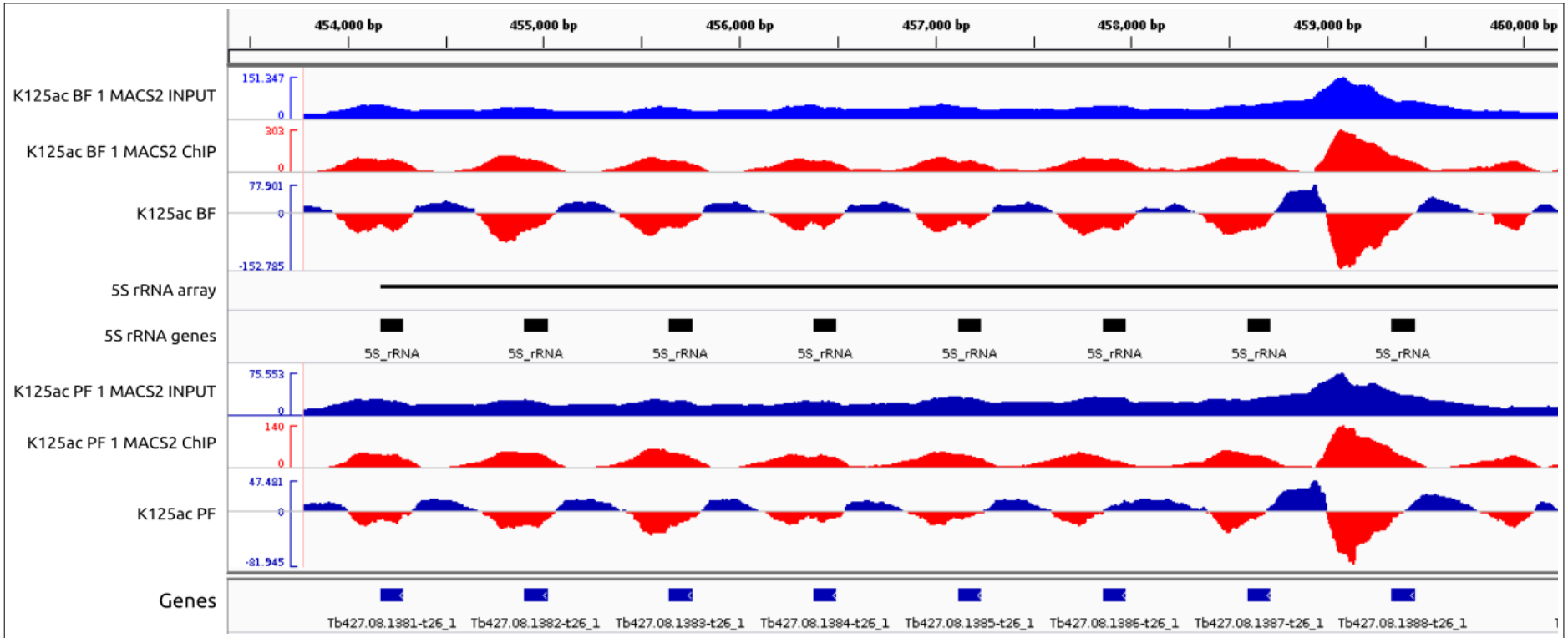


Figure 4.8. H2AK125ac samples displaying periodic nucleosomal patterns across the 5S rRNA gene array: Greater levels of non-K125ac nucleosomes (blue tracks) present between the 5S genes, and H2AK125ac containing nucleosomes (red tracks) on the 5S genes themselves. Samples 17 (BF, top tracks) and 20 (PF, bottom tracks) are shown as BF and PF representatives, with the 5S rRNA array and individual genes indicated in black.

Pol I transcribed rRNAs

The remainder of the ribosomal RNAs (i.e. 18S, 28S, 5.8S, large and small rRNA sub-units) transcribed by pol I, are interspersed throughout the MBCs. Intersection of these rRNAs with called peaks showed no correlation between enriched regions and the rRNA genes themselves (data not shown).

tRNAs

The *T. brucei* genome encodes 64 tRNAS distributed among the MBCs. Intersection of called peaks with tRNA genes revealed a remarkable enrichment of H2AK125ac marks present at tRNA genes (Fig. 4.9).

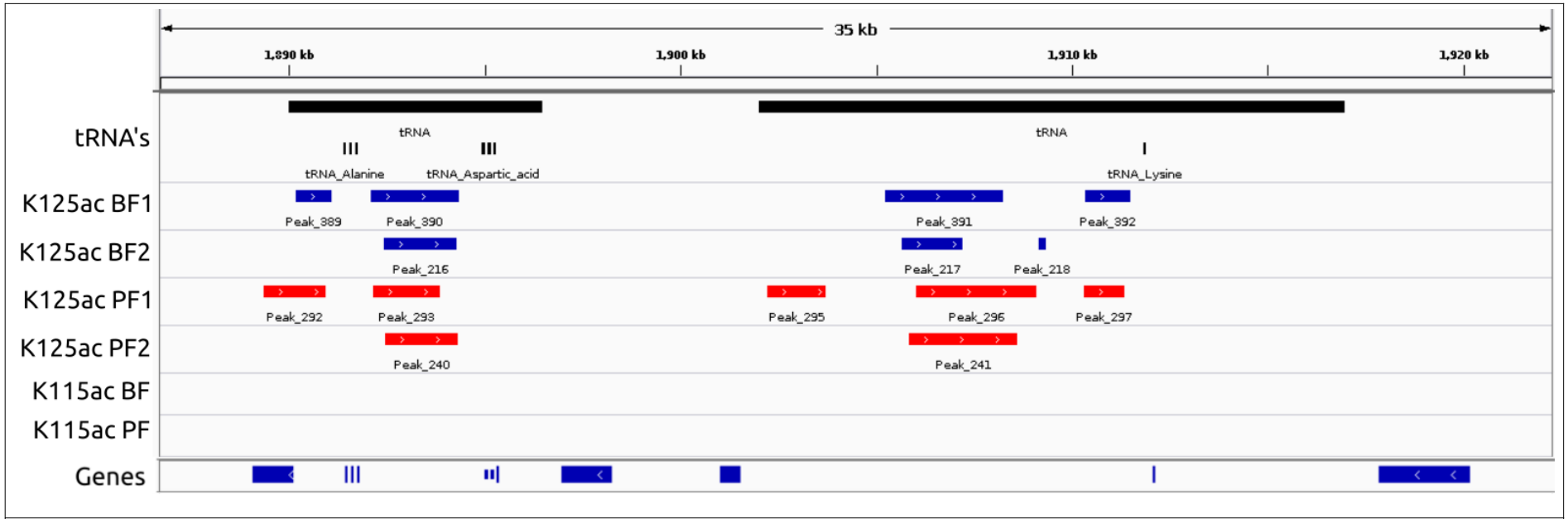


Figure 4.9. H2AK125ac peaks enriched at tRNA genes: Chromosome 7 (position 1888284 - 1918926) shows tRNA gene span shown by black bar, individual genes shown by small black bars, BF K125ac samples in blue, and PF in red. H2AK115ac samples show no enrichment at these sites.

H2AK125ac peaks present at tRNA loci were found to coincide with regions of transcription termination. As actively transcribed tRNAs will present a steric roadblock to a transcribing pol II, tRNAs have been found to be bordered by other epigenetic marks indicative of pol II transcription initiation, such as H2A.V and H4K10ac (Siegel et al. 2009). These tRNAs often coincide with pol II roadblocks, or internal pol II stop/start sites where putative pol II transcription initiation occurs (Fig. 4.10, A).

SL RNA

The spliced leader (SL) RNAs are transcribed at very high rates by pol II from monocistronic arrays located on chromosome 9, situated between two head-to-tail pol II PTUs (Daniels, Gull, and Wickstead 2010).

H2AK125ac peaks were found to be highly enriched at this loci, and K125ac peaks were observed covering the entire ~42.8 kb span of the SL RNA arrays (Fig. 4.10, B). The presence of hyperacetylated H2A at loci where pol II transcription occur at rapid rates suggests an intriguing correlation between H2A K125ac cPTM deposition and pol II mediated transcription.

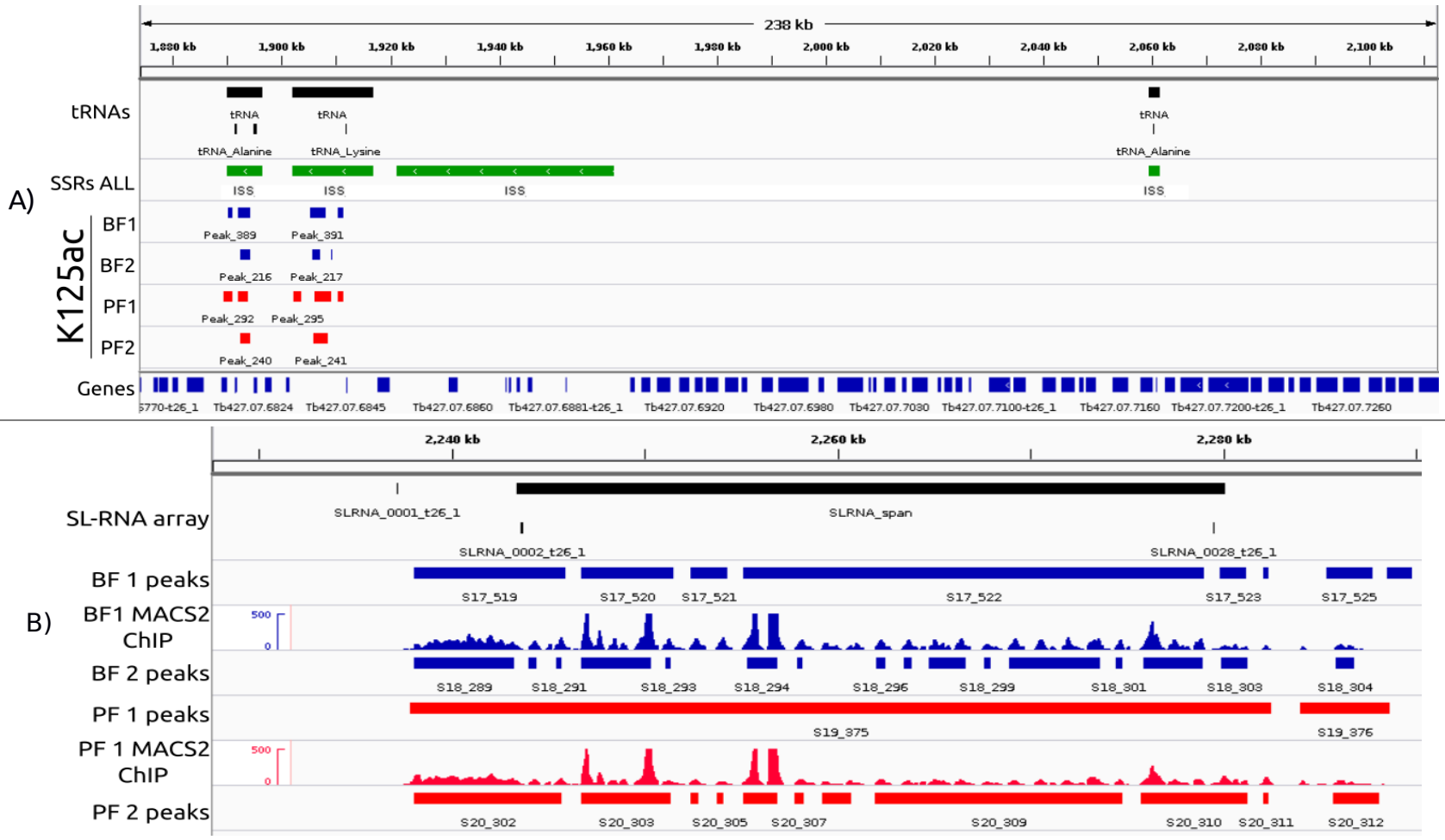


Figure 4.10: (**A**) Some tRNA genes are not bordered by H2AK125ac peaks. tRNA genes on chromosome 7 (position 1873913 - 2113887) with the tRNAs (same as shown in figure 4.9) on left hand side in panel A, flanked by K125ac peaks (~1880 - 1920 kb), but tRNA on right hand side (~2060 kb) devoid of K125ac peaks. (**B**) H2AK125ac peaks at pol II transcribed SL RNA genes in BF (blue) and PF (red) replicates. Nucleosomal ChIP profiles for BF (BF1 MACS2 CHIP) and PF (PF2 MACS2 CHIP) samples show regular peaks of enrichment, indicative of the repetitive arrangement of the SL RNA gene array.

Other genomic features

Other genomic loci, including polyadenylation sites, convergent SSRs, snoRNAs, life-cycle specific genes (like procyclin and PAG genes), and internal splice acceptor sites displayed no enrichment for the H2AK125ac or K115ac cPTMs. However, some SAS did intersect with H2AK125ac peaks. As the SAS is only 2 nt in length, and it has previously been found that SASs are occupied by a well-positioned nucleosome with the SAS itself located near the nucleosomal dyad (Maree et al. 2017), 75 nt was added to the up- and downstream regions of the SAS coordinates to intersect peak and SAS coordinates. The SASs that did display H2AK125ac enrichment were located at the 5' side of pol II PTUs (i.e. the first SAS of the first gene in a PTU, Fig. 4.11), and coincided with regions of pol II transcription initiation.

Localization of K125ac peaks at the first SAS of a pol II PTU, enrichment at the SL RNA arrays, and presence at the flanking regions of pol II 'roadblocks' (5S rRNAs and tRNAs at internal stop/start sites, supplementary document R4) suggested that H2AK125ac containing nucleosomes were present and markedly enriched at regions of putative pol II transcription initiation.

Figure 4.11. H2AK125ac peak enrichment at the first SASs of pol II PTUs. (**A**) A region of chromosome 7 (656662 - 106366) is shown as representative, displaying H2AK125ac peaks enriched at the most upstream SAS of PTUs in both BF (darker blue) and PF (darker red) replicates. These regions coincide with putative pol II TSSs, displayed in the 'SSRs' track as dSSRs (blue) and internal stop/start regions (ISS, green) as indicated in black boxes. No enrichment is observed at regions of putative pol II transcriptional termination (cSSRs, red). (**B**) K125ac peaks present at the most upstream SAS at the 1st gene of divergent PTUs. A representative region on chromosome 7 (position 1743 - 1755 kb) is shown; dSSRs and SASs are indicated by light blue and black bars, respectively. All K125ac peaks are shown next (BF dark blue, PF red) followed by intersected peaks for all K125ac and K115ac samples. No peaks are present in the K115ac samples.

4.4) Regions of pol II transcription initiation and termination

An unusual feature of trypanosomal pol II gene transcription is the absence of conventional eukaryotic pol II promoter and other regulatory sequences (Maree and Patterton 2014). With the exception of the SL RNA genes which contain the only well-defined trypanosomal pol II promoter, transcription initiates bi-directionally from divergent SSRs (dSSRs) as well as several HT regions (Martínez-Calvillo et al. 2010). Interruption of a pol II PTU by an RNA gene (like tRNAs, rRNAs, etc.) causes pol II transcription to terminate and re-initiate downstream of the RNA roadblock. These sites are known as internal stop/start sites (ISSs).

Divergent Strand Switching Regions:

. Intersection of called peaks with the above-mentioned strand switching regions unveiled a striking enrichment of H2AK125ac peaks at regions of putative pol II transcription initiation. Divergent SSRs were found to be highly enriched with H2AK125ac peaks throughout all 11 MBCs. All of the 74 annotated dSSRs examined were enriched with K125ac peaks in both BF and PF replicates (Fig. 4.12). Little to no enrichment is seen in either BF or PF H2AK115ac samplesNucleosomal profiles of ChIP and Input samples displayed a striking enrichment of K125ac peaks. H2AK125ac peaks can be observed across entire divergent strand switching regions, with K125ac nucleosomes being ~5x times greater than the background (input).

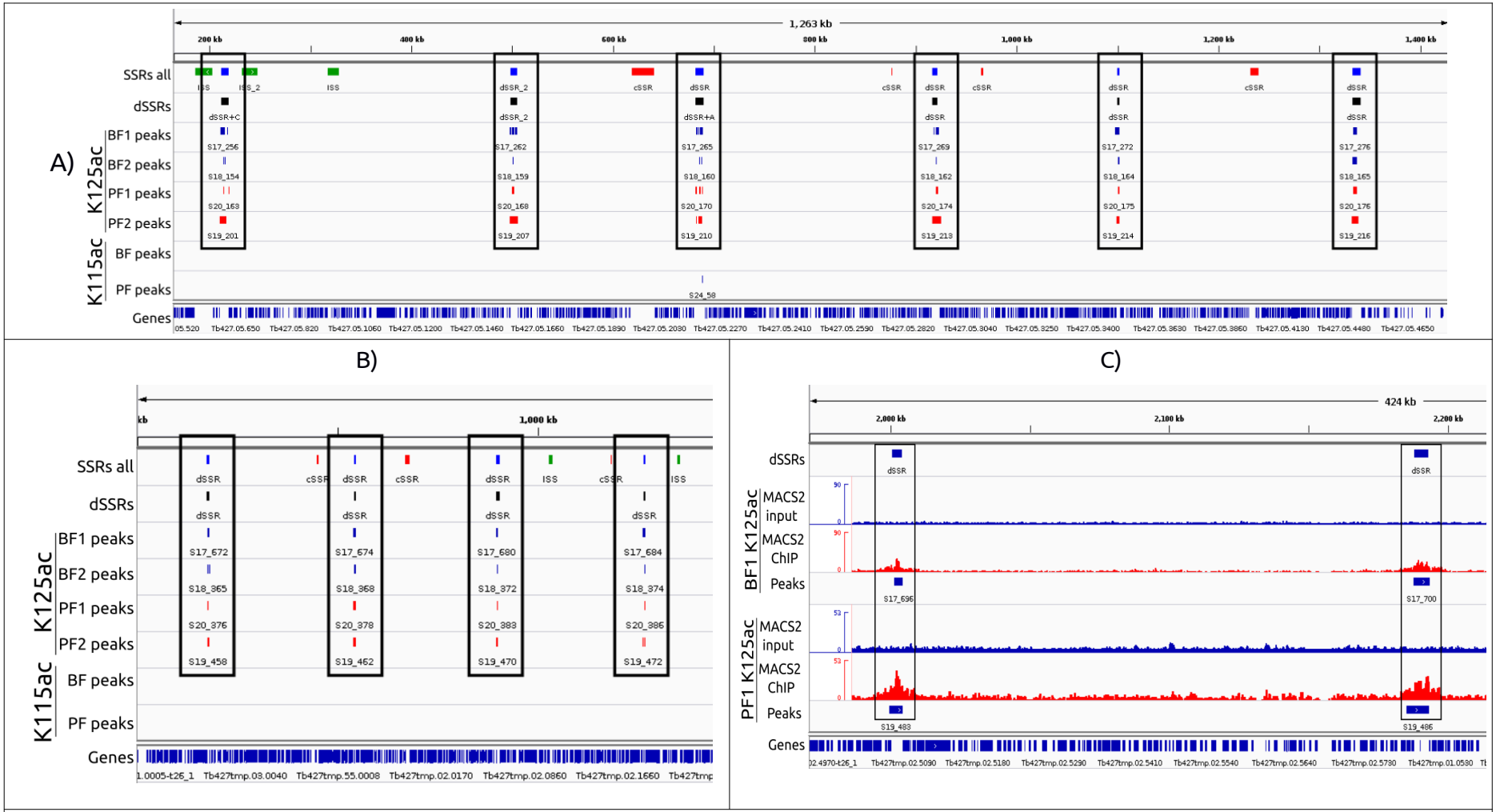


Figure 4.12. Histone H2AK125ac is enriched at dSSRs: Localization of K125ac peaks at divergent SSRs on (**A**) chromosome 5 and (**B**) chromosome 11. The top track shows all dSSRs (blue), ISSs (green) and cSSRs (red). Black boxes show the K125ac peak enriched at dSSRs for all K125ac BF (blue) and PF (red) samples. (**C**) Input (blue) and ChIP (red) profiles of BF and PF K125ac samples showing enrichment of K125ac containing nucleosomes at dSSRs.

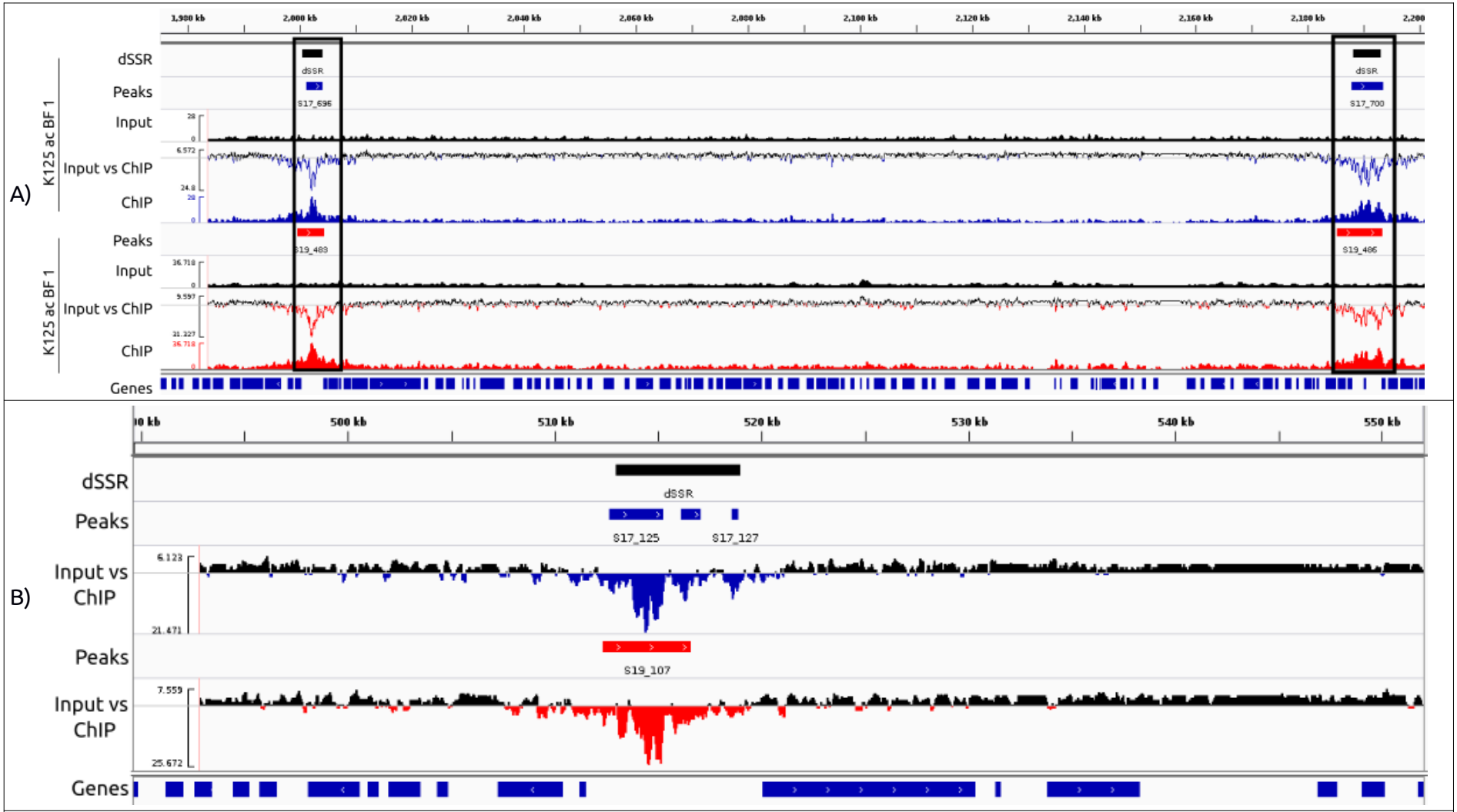


Figure 4.13. Nucleosomal profiles showing K125ac enrichment at dSSRs: Nucleosome profiles for BF and PF samples showing the input DNA (INPUT; black) and ChIP DNA samples (BF CHIP; blue | PF CHIP; red). (**A**) Line graphs of INPUT and CHIP profiles juxtaposed show enrichment of ChIP DNA (below x-axis) over input DNA (above x-axis) for BF and PF samples. dSSRs and K125ac enrichment are indicated by the black boxes. (**B**) Closeup of K125ac peaks at a dSSR, with INPUT juxtaposed with ChIP, BF in blue, PF in red.

Figure 4.13 (panel A) shows BF and PF K125ac Input and ChIP profiles for two dSSRs on chromosome 11; B shows a closer view of Input (black, above horizontal axes) vs ChIP (colored, below horizontal axes) for a dSSR on chromosome 2. K125ac peaks present at dSSRs are localized to the beginning of the pol II PTU, and does not significantly extend further downstream into the PTU. This indicates that K125ac marks are localized at pol II transcription start site at dSSRs.

[*pol II PTU] Internal Stop/Start sites:*

H2AK125ac peaks were observed to be present at regions where pol II transcription terminates and re-initiate at HT and ISS regions. This may be caused by the presence of an actively transcribed tRNA gene which presents a kinetic block to pol II, and contributes to its termination. Of all of the 62 ISSs investigated, 95% (n=59) showed K125ac enrichment (Fig. 4.14, A). Nucleosome profiles of Input vs ChIP DNA show a clear enrichment of K125ac at ISS (Fig. 4.14, B). Some of the ISS loci enriched with K125ac peaks coincided with tRNAs, known regions of pol II stop/start sites (Fig. 4.14, C). K125ac nucleosomes observed at these loci were located downstream of the tRNA genes, extending somewhat downstream into the PTU.

Convergent Strand Switching Regions:

Of the 45 cSSRs investigated, most was found to be devoid of any PTM investigated. The cSSRs that did intersect with enriched peaks often coincided with tRNAs, centromeric regions, or regions containing genes with known high copy numbers (Fig 4.15).

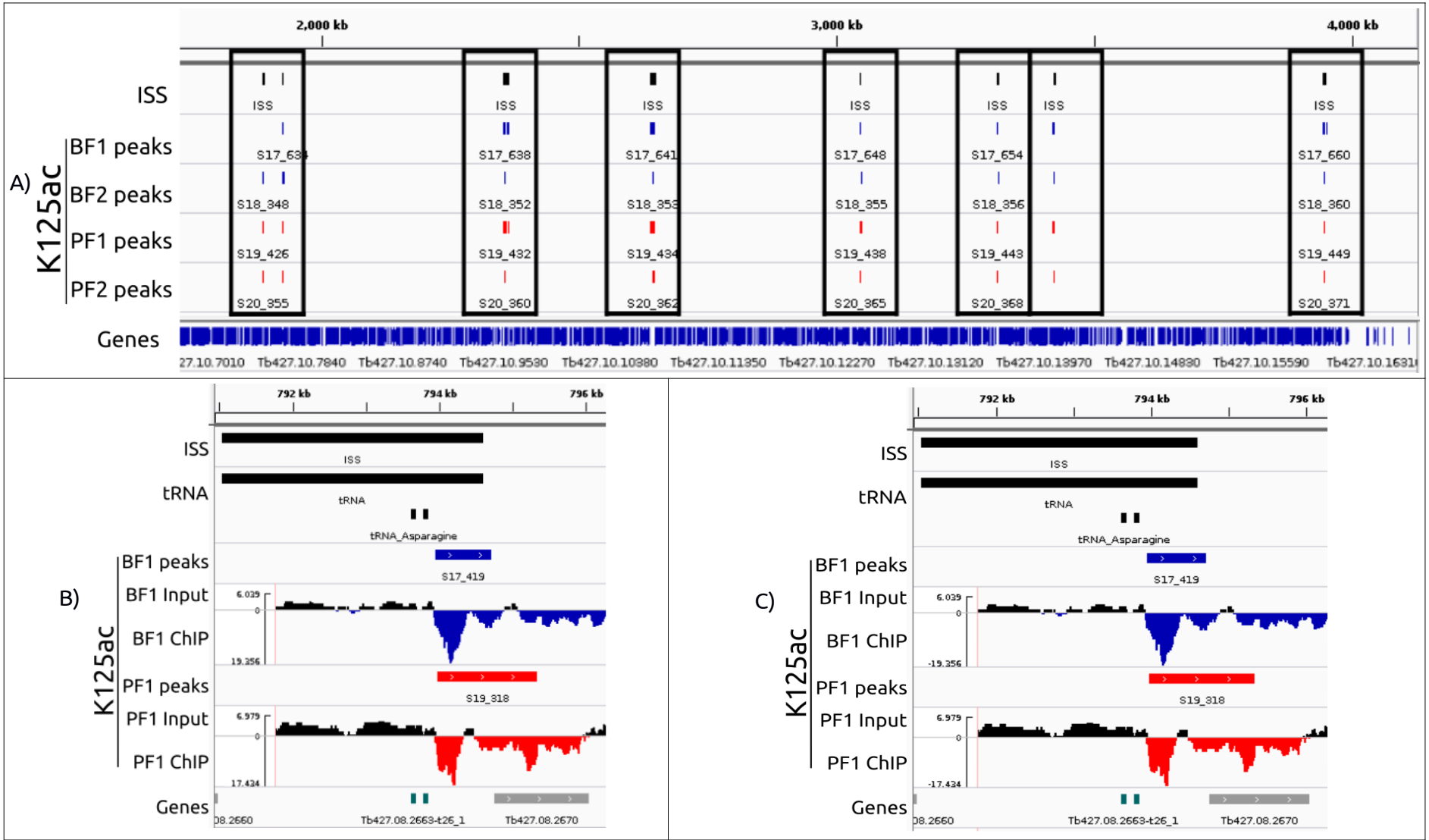


Figure 4.14. Histone H2AK125ac is enriched at ISS: (**A**) BF (blue) and PF (red) K125ac ChIP samples show peak enrichment at all internal pol II stop/start sites (black bars) on chromosome 10. ISSs and regions of K125ac enrichment are indicated by black boxes. (**B**) Nucleosome profiles for BF and PF samples showing input DNA (INPUT; black) juxtaposed with ChIP DNA samples (17 CHIP; blue | 19 CHIP; red). Line graphs between INPUT and CHIP profiles show enrichment of ChIP DNA (below x-axis) over input (above x-axis) for BF (17 BF K125ac) and PF (19 PF K125ac) samples. (**C**) K125ac peaks at tRNA genes.

Figure 4.15. ChIP peaks of enrichment called at convergent SSRs: (**A**) cSSR on left-hand side of chrom 1 (position 730 kb) devoid of any ChIP enrichment. cSSR on the right-hand side (780 - 800 kb) coincides with the centromere (black), and shows ChIP enrichment across majority of samples. (**B**) cSSR enriched for K125ac peaks on chrom 8 downstream with tRNAs (green). (**C**) called peaks at a cSSRs on chromosome 4, bordered by genes with known high copy numbers (yellow = LRRP and RHS genes).

Taken together, the observed co-localization of H2AK125ac peaks at regions of putative pol II transcription initiation (at dSSRs, ISSs, the 1st SAS of a PTU, pol III transcribed RNAs, and at the SL RNA array), and absence at termination regions (cSSRs) suggests that the K125ac PTM pattern marks areas of putative pol II transcription start sites (TSSs).

It was previously observed that histone variants H2A.V and H2B.V, and RPB9 (a pol II sub-unit) coincide with pol II transcription start regions at dSSRs and ISSs (Wedel et al. 2017). As *T. brucei* depends on the epigenome to delineate its genome and regulate gene function, it would be of interest to determine whether the observed K125ac cPTM enrichment co-localized with the aforementioned marks indicative of pol II TSSs. Therefore, MNase-seq data from a recent study were retrieved from the GEO database (accession number GSE98061, Wedel et al. 2017) and processed with the same data processing pipeline used for the ChIP-seq data to identify regions of H2A.V and RPB9 enrichment. As previously observed, twin peaks of histone H2A.V were seen flanking dSSR - indicative of bidirectional pol II transcription (Fig 4.16, A). A single H2A.V peak is present on ISSs, and occurs downstream of tRNA genes when present at ISSs (Fig 4.16, B). RPB9 displayed analogous peak enrichment and profiles to that observed for H2A.V at both dSSRs and ISSs (Fig. 4.17, A and B). Both H2A.V and RPB9 ChIPs display the greatest enrichment in the immediate proximity of pol II TSSs, but can also be detected at basal levels throughout the PTU (Supplementary document R5).

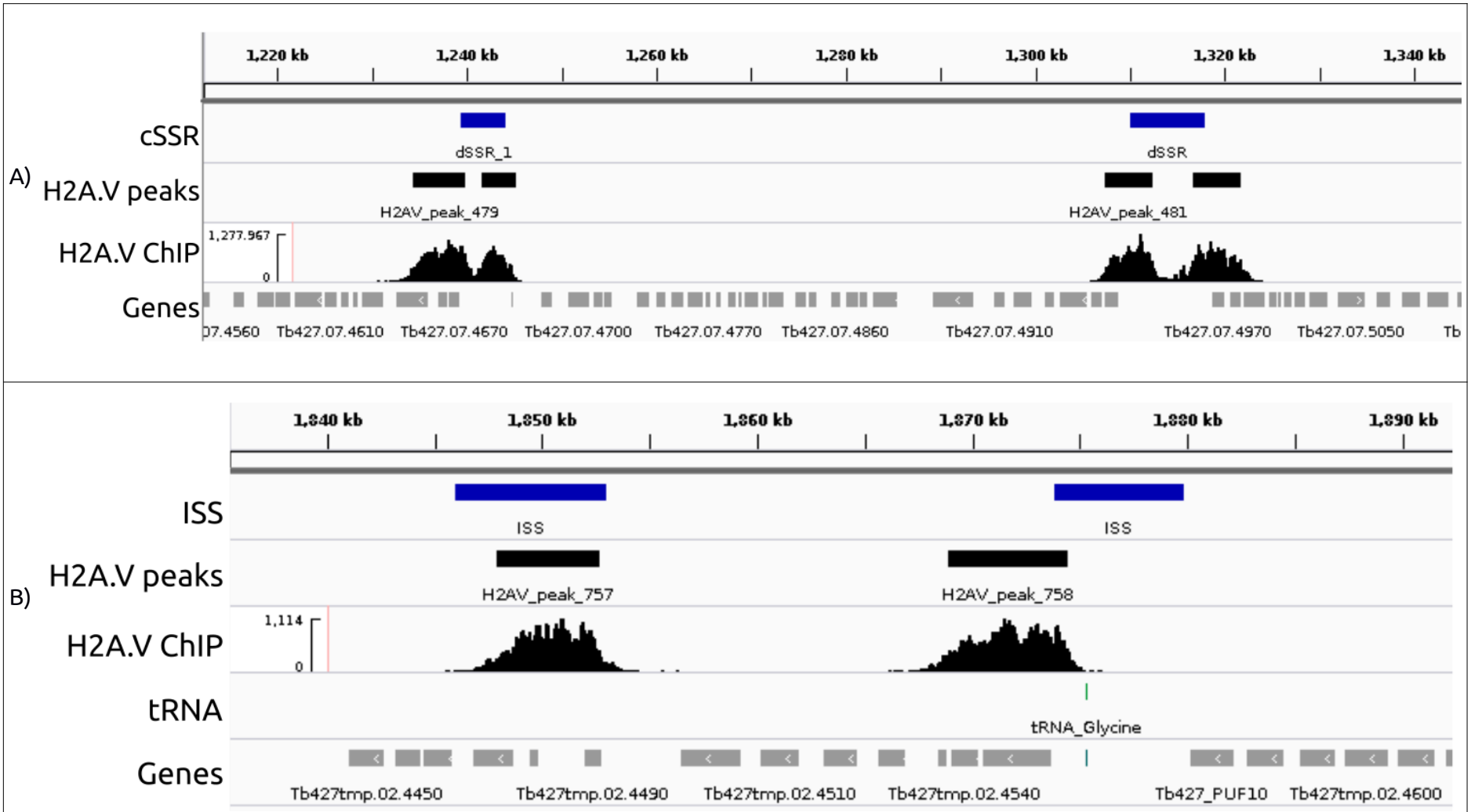


Figure 4.16. Histone variant H2A.V at putative pol II TSS: (**A**) Twin H2A.V peaks (black boxes) can be seen at two dSSRs (blue) on chromosome 7 (positions 1215354 - 1353925), with the black histograms showing H2A.V nucleosome positions. (**B**) Single H2A.V peaks and nucleosome profiles at ISSs on chromosome 11 (positions 1834269 - 1897595). Genes are indicated by gray boxes, tRNA genes in green. Shown loci is representative of all MBCs.

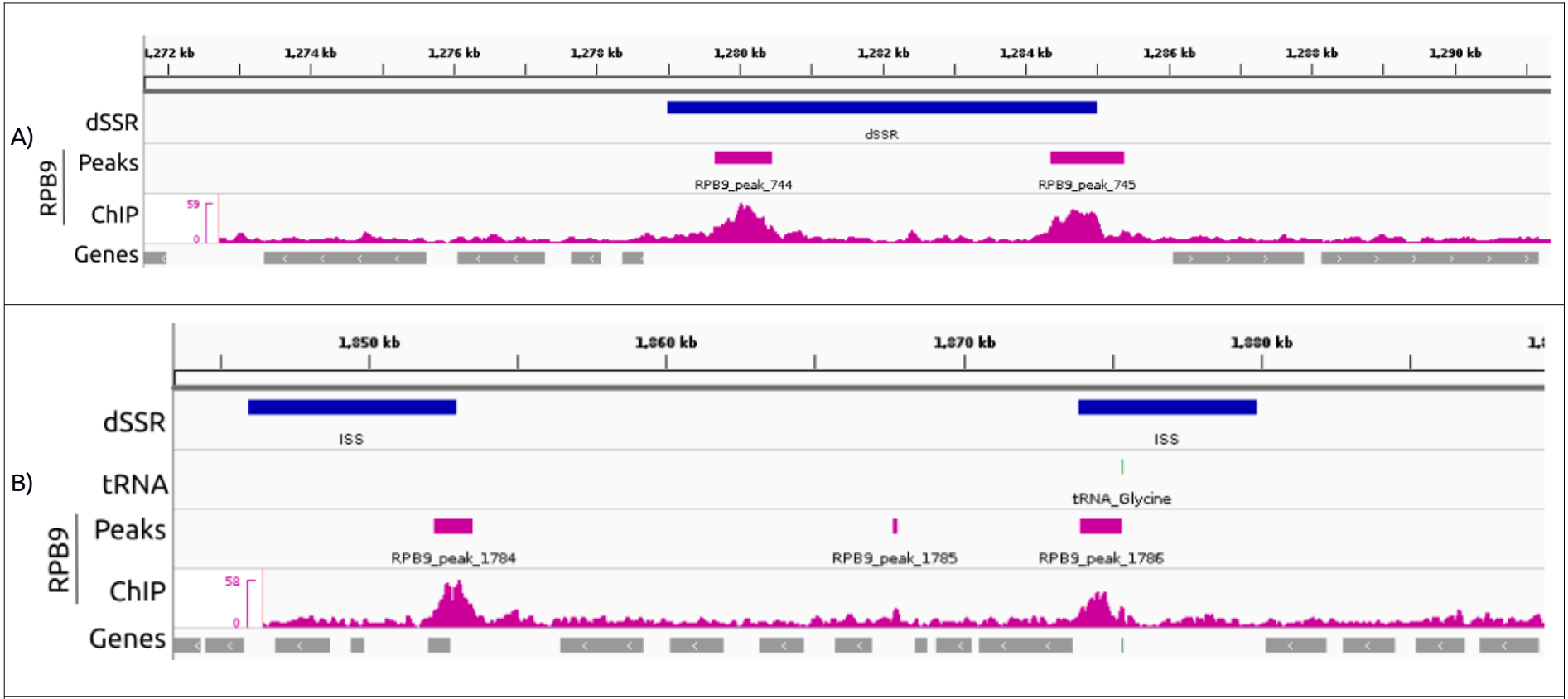


Figure 4.17. RPB9 at putative pol II TSS: (**A**) Twin RPB9 peaks (magenta boxes) can be seen at two ISSs (blue boxes) on chromosome 6 (positions 1271663 - 1292214), with the black histograms showing RPB9, in essence pol II positions. (**B**) Single RPB9 peaks and pol II positioning at ISSs on chromosome 11 (positions 1834269 - 1897595). Genes are indicated by gray boxes, tRNA genes in green.

The presence of pol II at low levels across entire pol II PTUs was expected as the housekeeping portion of the *T. brucei* genome is constitutively transcribed. Pol II is, therefore, likely to be present along the entire pol II polycistronic transcription unit. When compared, H2A.V and RPB9 display remarkably similar distribution profiles at dSSR and ISS regions (Fig. 4.18), and serve as useful indicators of pol II TSSs in trypanosomes. For convenience, the same genomic loci were used in figures 4.16 - 4.18 to facilitate comparison between figures.

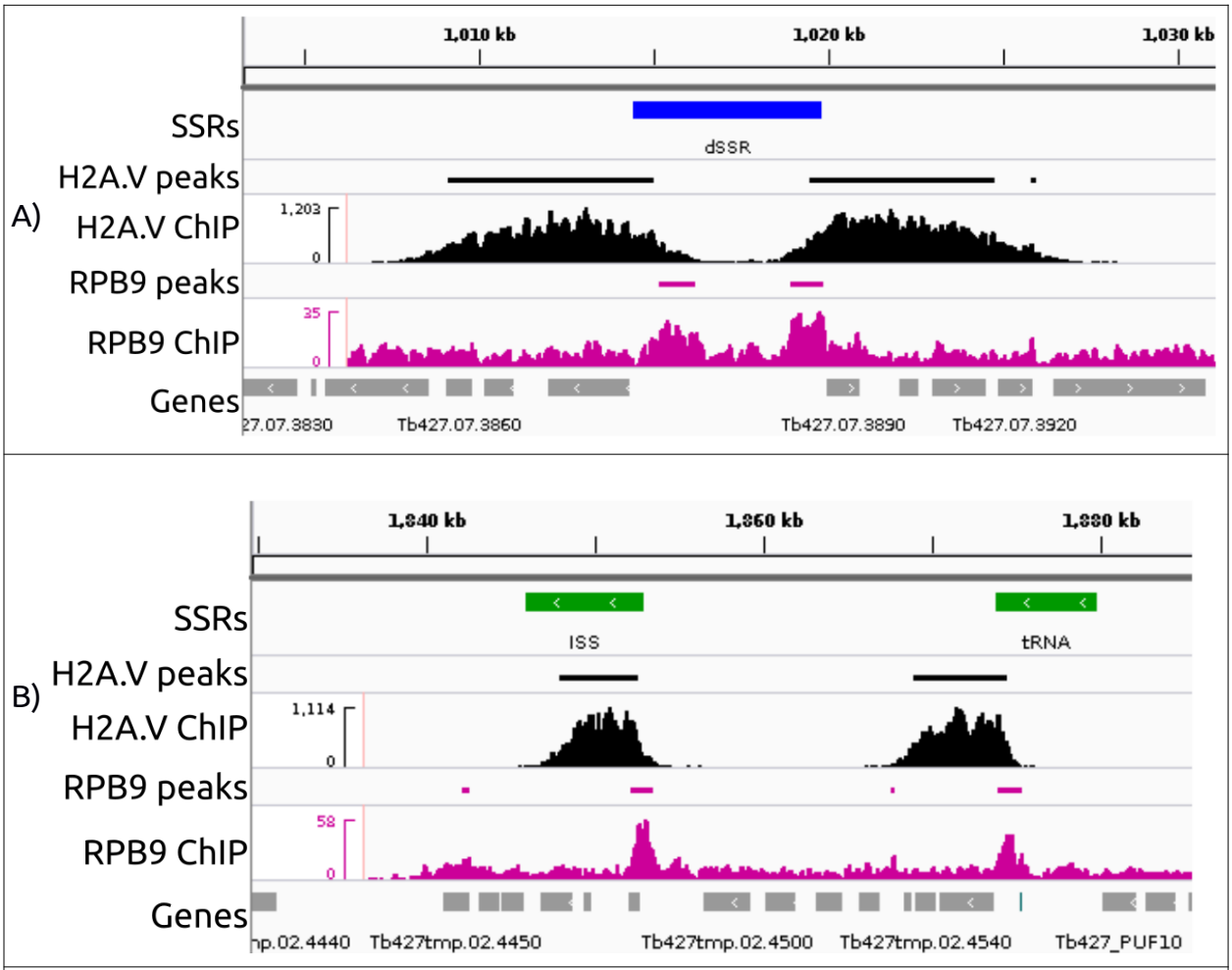


Figure 4.18. H2A.V and RPB9 peaks co-localize at putative pol II TSS: H2A.V (in black) and RPB9 (magenta) displaying co-localized enrichment profiles at (**A**) dSSRs (in blue) and (**B**) ISSs (in green).

Comparing the loci enriched with both H2A.V and RPB9 peaks to those enriched with K125ac peaks, a striking correlation between these enriched loci and areas of putative TSSs was visible. H2A.V and RPB9 peaks coincided with K125ac peaks at all 74 dSSRs investigated, and was present in both BF and PF replicates (Fig 4.19).

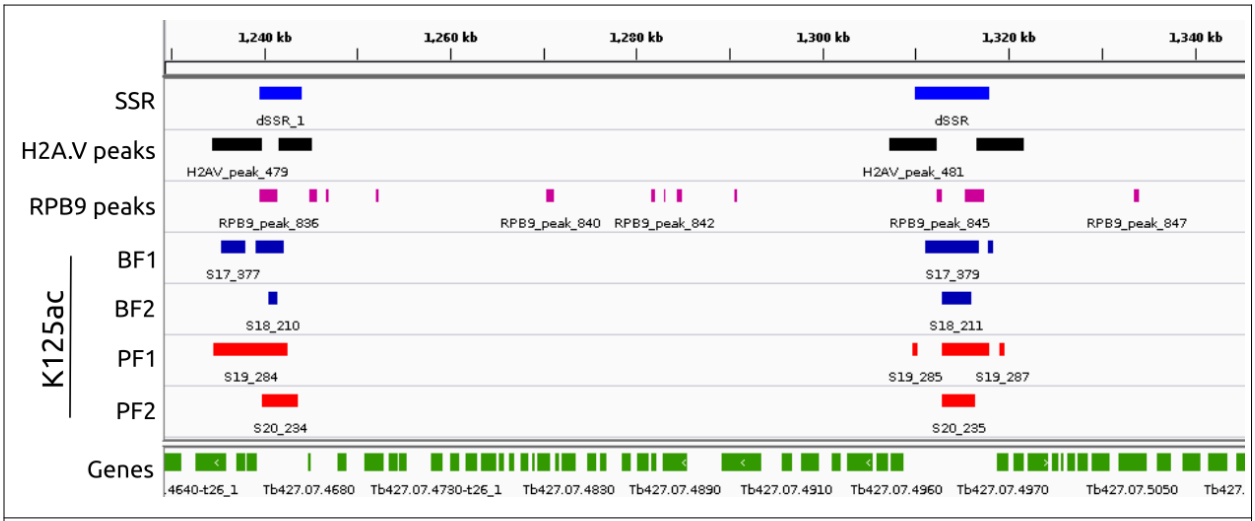


Figure 4.19. H2A.V, RPB9, and K125ac peaks co-localize at dSSRs: A representative region of chromosome 7 (position 1209978 - 1339924) showing peak enrichment of histone variant H2A.V (black), RPB9 (magenta) and K125ac samples (BF replicates in dark blue; PF replicates in red) colocalizing at divergent strand switching regions (light blue). Genes are shown in green.

ChIP profiles showing nucleosome positions containing the K125ac cPTM pattern revealed that K125ac marks are highly enriched in between the twin H2A.V peaks and the immediate vicinity of dSSRs, and extended downstream into the PTU to basal levels (Fig. 4.20).

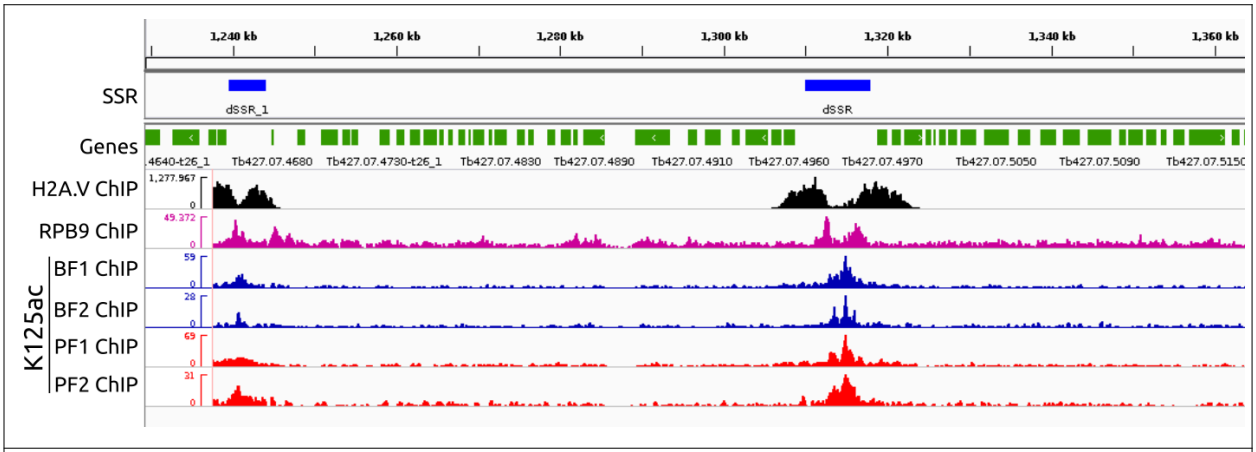


Figure 4.20. H2A.V, RPB9, and K125ac ChIP profiles at dSSRs: Chromosome 7 (position 1209978 - 1339924, same region as shown in fig. 4.19) displaying ChIP profiles of H2A.V (black), RPB9 (magenta), K125ac samples (BF's in dark blue; PF's in red). H2AK125ac peaks are present between twin peaks of H2A.V.

Internal stop/start regions also displayed similar localization of histone H2A.V, RPB9, and K125ac peaks. At these HT sites, RPB9 and K125ac peaks are present on the ISS site, with a single H2A.V peak present immediately downstream of the ISS (Fig 4.21). ChIP profiles revealed that K125ac nucleosomes do not extend bi-directionally as observed at dSSRs, but propagates to the downstream side towards the 5' side of the PTU, decreasing to basal levels similar to that seen at dSSRs. Supplementary document R6A and B shows additional data of K125ac peaks and ChIP profiles coinciding with putative pol II transcription start regions at dSSRs and ISS sites. Taken together, the data strongly suggests that *T. brucei* utilizes a combinatorial PTM pattern to selectively hyperacetylate H2A N-terminal tails at discrete loci of known pol II transcription initiation. The H2AK125ac combinatorial PTM, in addition to other epigenetic marks like H3K4me3, H4K10ac, and histone variant deposition, is present at regions where pol II transcriptional machinery assembles and transcription initiates (Mandava, Janzen, and Cross 2008;

Siegel et al. 2009; Wright, Siegel, and Cross 2010). These may all form part of an intricate histone code used by *T. brucei* to regulate a genome that lacks canonical pol II transcriptional control.

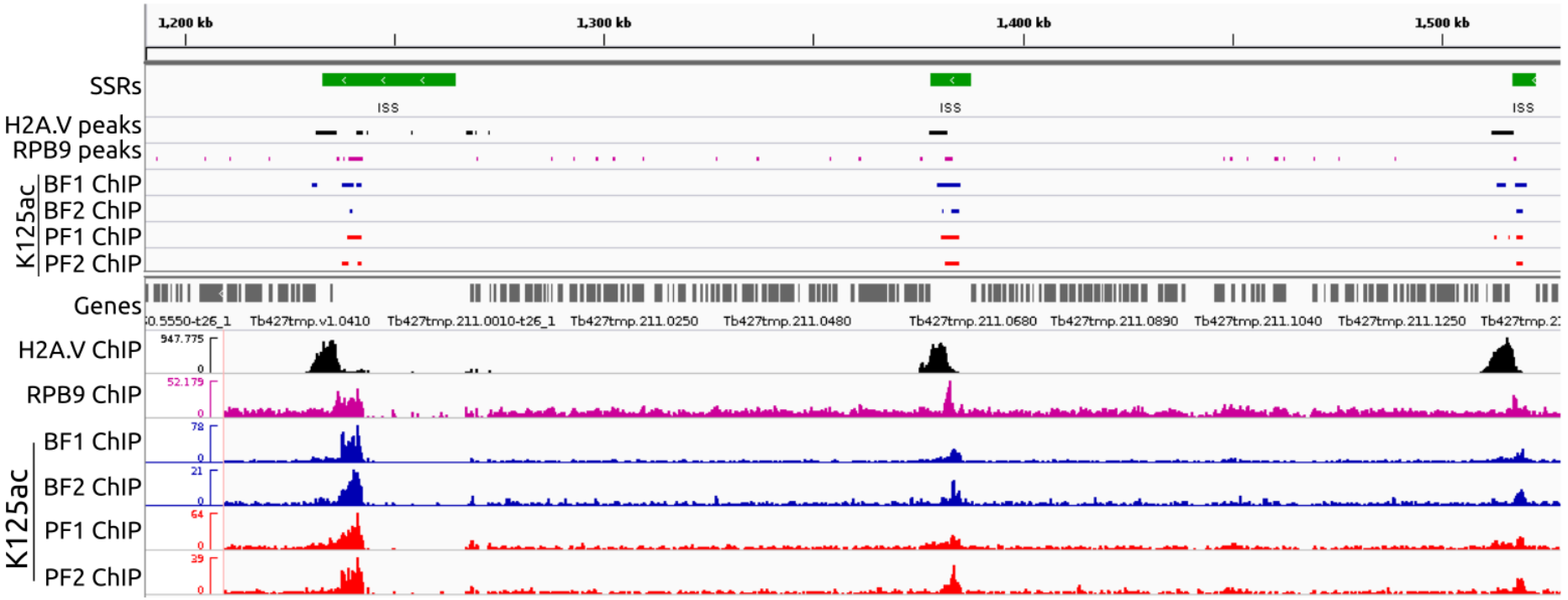


Figure 4.21. H2A.V, RPB9, and K125ac peaks and ChIP profiles at ISSs: A representative region on chromosome 9 (position 1190579 - 1557867) showing three internal stop/start sites of head-to-tail PTUs. Single H2A.V peaks (black) can be seen downstream of the ISS sites (green), with RPB9 (magenta), BF replicates (dark blue), and PF replicates (red) replicates localizing on the ISSs. Genes are indicated in gray. PTU directionality is indicated by the arrows in the green ISS blocks.

5.0 Investigation of the nuclear distribution of hyperacetylated H2A by Immunofluorescence microscopy.

Investigation of the eukaryotic nuclear architectural configuration by immunofluorescence microscopy (IF) has shed light on how the nucleus is spatially and functionally partitioned. The nuclear pore complex (NPC) is a large multi-protein complex that facilitates (macro)molecular exchange between the nucleus and cytoplasm (Wente and Rout 2010; Ersfeld 2011). Proteomic investigation of the NPC in *T. brucei* revealed a high degree of conservation of NPC components, suggesting functionality akin to that seen in higher eukaryotes (Ersfeld 2011). The inner nuclear membrane in metazoan cells is lined with lamin proteins - intermediate filament proteins that provides structural support to the nucleus. Conserved functions such as chromatin organization, gene and transcriptional regulation, as well as cell-cycle progression have been observed for *T.brucei* NPCs and lamin proteins (Dechat et al. 2010; Mekhail and Moazed 2010).

In addition, it was found that nuclear functions such as genome replication and gene transcription are not homogeneously distributed throughout the nucleus, and occur in compartmentalized foci (Daniels, Gull, and Wickstead 2010). BrdUTP labeling of permeabilised cells revealed that trypanosomes, despite the early divergence from the main eukaryotic lineage, also compartmentalized replication and transcriptional activities.

Transcription by pol I, II, and III was observed to localize at distinct areas in the nucleus, showing discrete foci, some of which were localized at the nuclear periphery (Navarro and Gull 2001; Navarro, Peñate, and Landeira 2007). In addition to the nucleolus, the nuclear compartment where pol I mediated rRNA transcription occurs, a functionally distinct trypanosome specific nuclear compartment known as the expression site body (ESB) can be observed. The ESB is crucial for VSG transcription and mono-allelic exclusion in BF cells and contains the pol I transcribed bloodstream expression site where the active VSG is polycistronically transcribed along with ESAG genes (Navarro and Gull 2001; Daniels, Gull, and Wickstead 2010). IF microscopy is a powerful technique that allows visualization and investigation of nuclear compartmentalization. However, it should be noted that IF microscopy only provides a snapshot of the genes being actively transcribed at that *specific* time point.

IF microscopy revealed that transcription by pol II in trypanosomes was present at several foci and was seen to occur throughout the entire nucleus (Smith et al. 2009). As H2AK125ac was observed to localize to putative pol II TSS, it would be of interest to determine whether this cPTM displayed similar nuclear foci to that of pol II transcription. The H2AK125ac and H2AK115ac antibodies were used to examine their nuclear distribution by IF microscopy. ChIP grade antibodies can be used to target specific antigens, such as histone PTMs, after which a secondary antibody attached to a fluorescent dye is used to visualize the distribution of specific epitopes. IF microscopy revealed that the two C-terminal hyperacetylation patterns on histone H2A, K125ac and K115ac,

displayed different localization foci. K125ac was seen throughout the entire nucleus, with foci appearing at the nuclear periphery and internal regions in both BF and PF cells (Fig. 5.1 A and B, respectively).

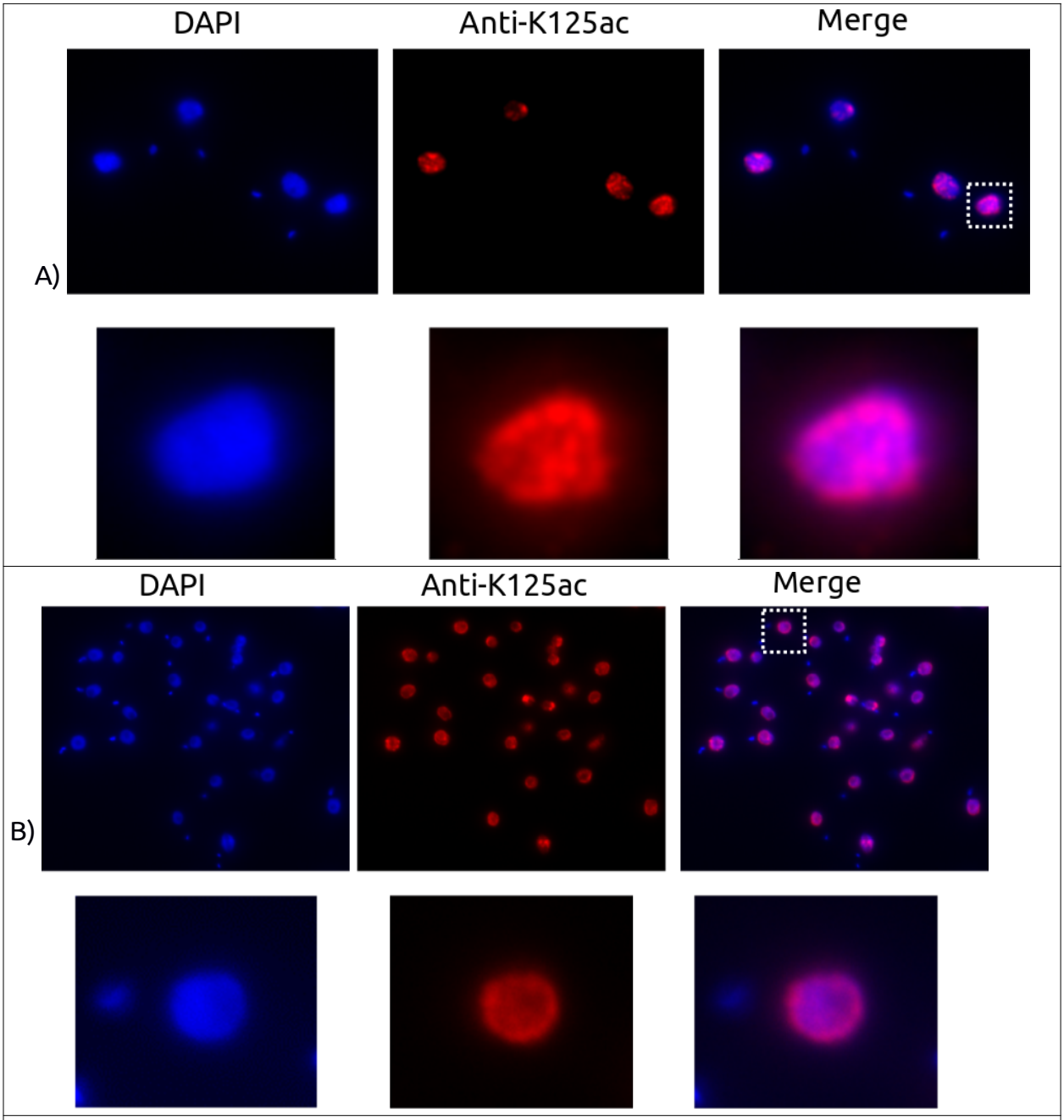


Figure 5.1: IF microscopy of H2AK125ac in (**A**) BF and (**B**) PF cells. Images show DAPI staining in blue, anti-K125ac in red and merge images.

Despite not showing an apparent genomic localization associated with specific gene classes or genomic boundaries, H2AK115ac was observed to be present mostly at the nuclear periphery in BF and PF cells (Fig. 5.2 A and B, respectively).

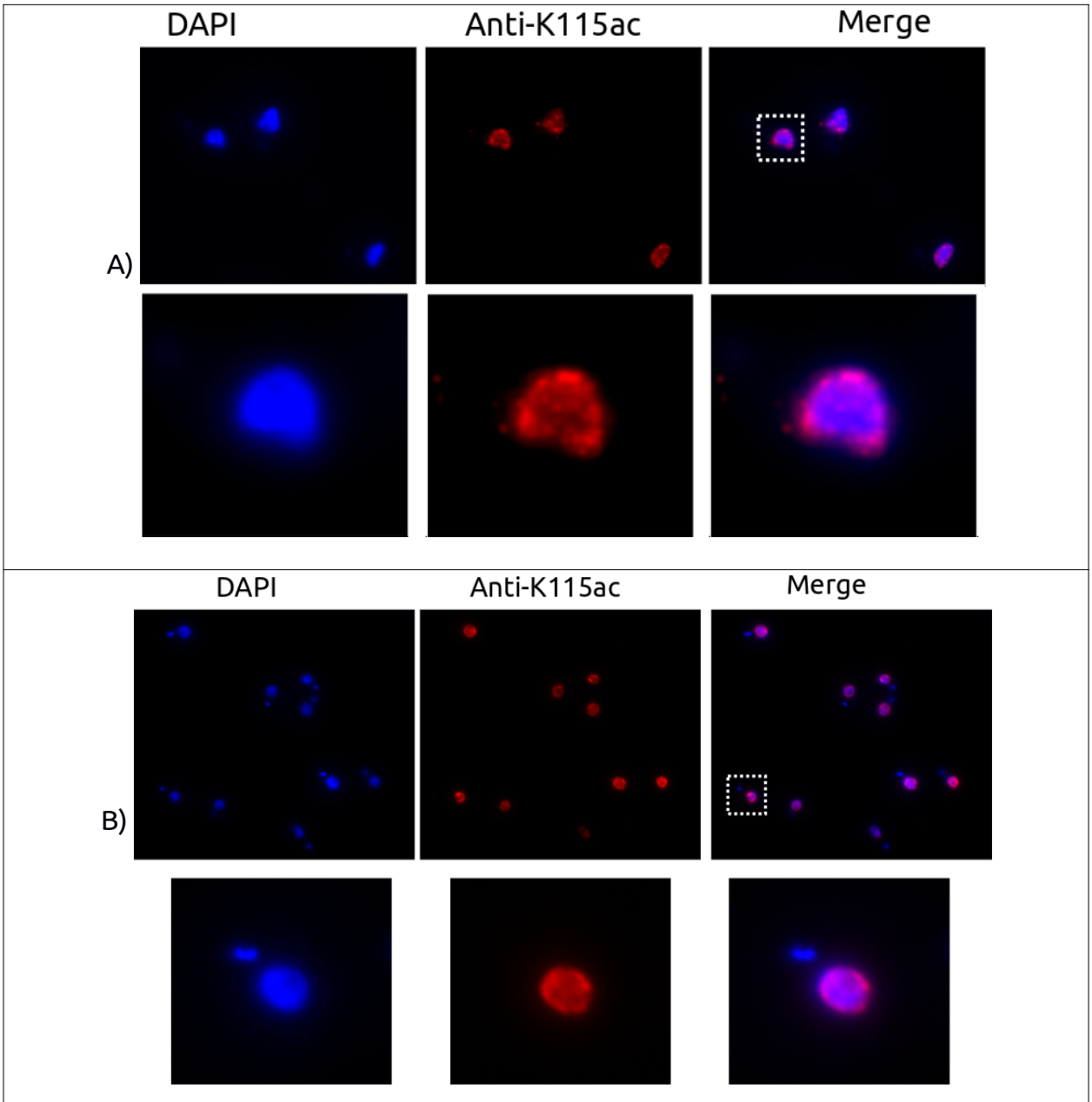


Figure 5.2: IF microscopy of H2AK115ac in (**A**) BF and (**B**) PF cells. Images show DAPI staining in blue, anti-K115ac in red, and merge images.

Figure 5.3 shows side by side comparison of H2A K125ac and K115ac IF images. Both K125ac and K115ac foci appeared more tightly focused in BF cells than PF cells, which may be a result of chromatin in the BF cells being more heterochromatic than in PF cells (Povelones et al. 2012). The foci displayed by H2AK125ac seems to be comparable to that previously seen for pol II, with discrete foci seen throughout the entire nucleus.

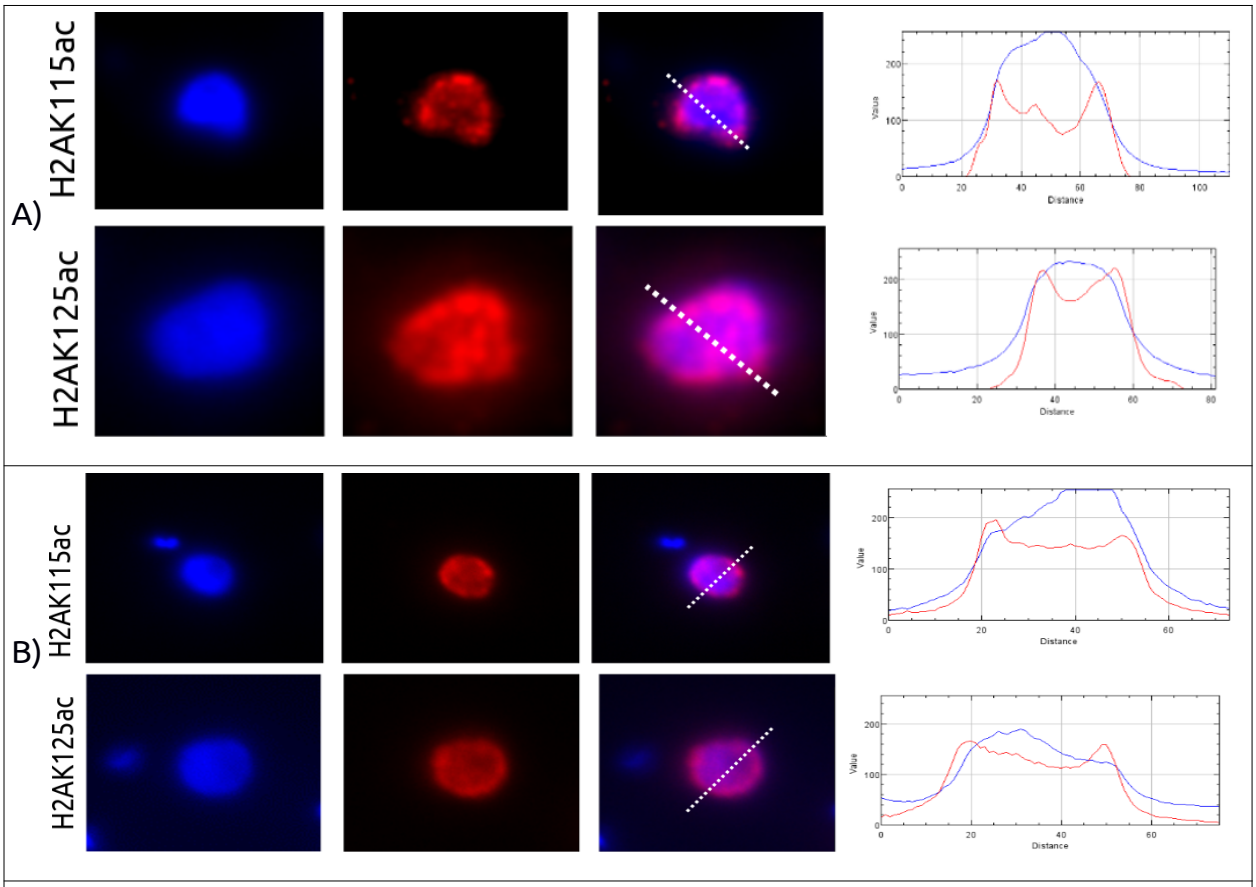


Figure 5.3: Comparison of H2AK115ac and H2AK125ac nuclear localization. (**A**) BF cells, and (**B**) PF cells showing DAPI (blue), active mark (red), merge, and z-stacks.

The K125ac distribution throughout the nucleus may be present at sites of active pol II transcription initiation regions as ChIP-seq data showed a clear correlation with known pol II TSS at dSSRs and ISS regions. Some K125ac foci seemed to form a ring around the nuclear periphery. The nuclear periphery can serve a transcriptionally repressive function (Finlan et al. 2008), however, actively transcribed genes have previously been observed at this location, with BrUTP labelling indicating the presence of active transcription foci (Navarro and Gull 2001; Navarro, Peñate, and Landeira 2007). K115ac did not display discrete foci like that seen for pol I foci such as the nucleolus and ESB. The association with the nuclear periphery is intriguing and may indicate that this mark is associated with gene repression, or it may serve an architectural function. As the telomere repeat sequences, intermediate-, and mini-chromosomes do not form part of the reference genome it is not clear if this mark is associated with these regions and chromosomes where it could be associated with repetitive 177bp sequences which are scarce on the megabase-chromosomes (Wickstead, Ersfeld, and Gull 2004). No obvious compartmentalization can be observed for either marks in BF or PF cells, or it may exist beyond the resolution of IF. Comparing life cycles, both K125ac and K115ac foci appear more focused in BF cells than in PF cells. This may be a result of a more relaxed and open chromatin structure in PF cells (Povelones et al. 2012), with K115ac foci appearing much more diffused in PF cells. If K115ac serves a transient repressive role this may be caused by stringent BES repression in BF cells not required in PF cells.

6) Discussion

Kinetoplasts diverged from the main eukaryotic lineage about ±1.5 billion years ago, a mere 300 million years after the first eukaryotes were believed to have emerged (Knoll et al. 2006; Cavalier-Smith 2010; He et al. 2014). This early branch-off gave rise to trypanosomes developing some unique mechanisms for adapting to complex host environments. *T. brucei* exhibits peculiar genomic characteristics like transcription of a major cell-surface protein by pol I, absence of canonical pol II promoters and other regulatory sequences, and polycistronic gene organization and transcription on a genome-wide scale. However, trypanosomes display intriguing correlations to epigenetic processes shared among higher eukaryotes, such as nucleosome positioning, utilization of histone PTMs and histone variants for genome demarcation, modulation of heterochromatic gene repression, and other epigenetic mechanisms that regulates genomic function (L. M. Figueiredo and Cross 2010; Povelones et al. 2012; Maree et al. 2017; Wedel et al. 2017). It has become increasingly clear that in *T. brucei*, chromatin and in particular the epigenome, have vital roles in genome regulation, and is essential for effective host immune evasion during the mammalian infective life stage (Wyse et al. 2013; Hughes et al. 2007; Rudenko 2010; Bañuelos et al. 2019).

Dynamic regulation of chromatin is often executed by post-translational modification of histone proteins. Complex patterns of combinatorial histone PTMs has recently been observed in a related kinetoplast, *T. cruzi* (De Jesus et al. 2016; Picchi et al. 2017).

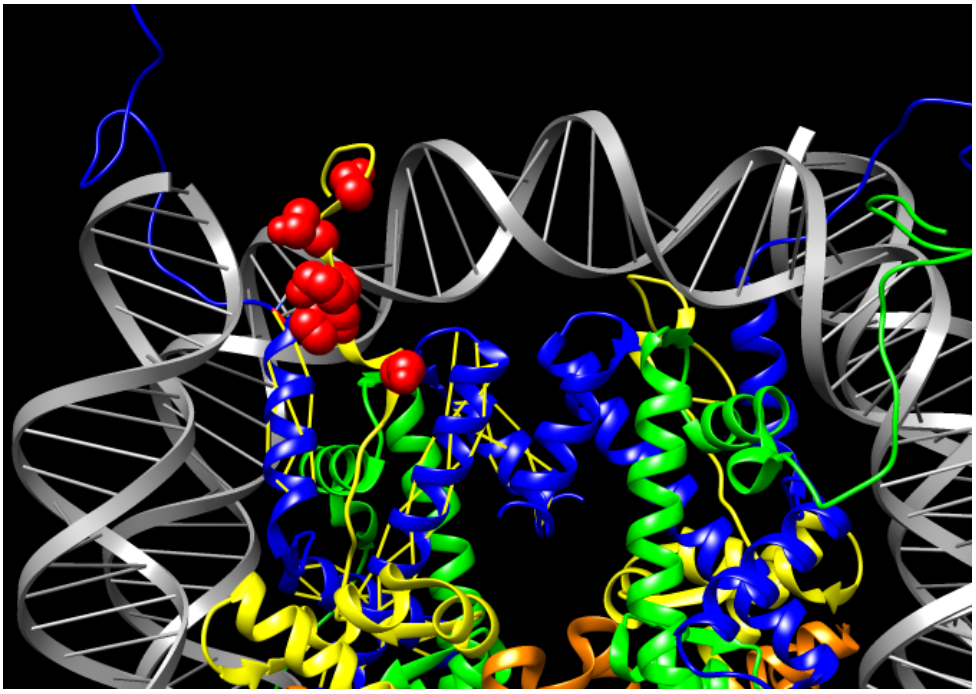
Numerous PTMs have been observed in *T. brucei*, with all four core histones displaying amino acid modifications. Despite divergence in histone amino acid sequences, some PTMs seem to be analogous to known PTMs in other eukaryotes (Maree and Patterton 2014). This study further explored the epigenetic repertoire of *T. brucei* histone PTMs by using a middle-down MS technique to ascertain combinatorial PTM patterns on both canonical and variant histones.

The N-terminal tail of H2A, like that of *T. cruzi,* appeared sparsely modified (Table 3.3). Previously observed mono-methylation of Ala1, and acetylation of Lys4 was seen. In addition, Lys9me1 and Ser11ac were detected. H2A proline 26 was one of two proline residues observed to be present as hydroxyproline and presents a novel PTM not previously observed in kinetoplasts.

In contrast to the N-terminal of H2A, the C-terminal was found to be heavily modified, with up to 5 acetylation marks found to be distributed between lysine residues K115, K119, K120, K122, K125, and K128 in various combinatorial patterns. Up to 50% of C-terminal H2A peptides were acetylated one or more times. This hyperacetylated state of the H2A C-terminal will negate the positive charge of the basic lysine residues, likely influencing DNA-octamer interaction by abrogating binding to the DNA duplex in this region, leading to a more euchromatic chromatin state (Fig. 6.1, A). This reduction in binding could have a structural effect similar to that of human H2A.Bbd (Fig 6.1, B). H2A.Bbd lacks the docking domain implicated in H1 and H4 binding, decreases protection of ~10 bp of DNA from each core terminus from enzymatic

digestion, is associated with a structurally destabilized nucleosome, and abolishes the ability of the H2A.Bbd containing chromatin to condense (Arimura et al. 2013). It has been established that an open, more relaxed chromatin state may lead to the precipitation of pol II transcriptional machinery and lead to fortuitous transcription initiation in *T. brucei*. Indeed, BF parasites have more heterochromatic chromatin than PF cells, which contributes to transcriptional silencing of VSG gene arrays (Povelones et al. 2012).

A)



B)

```
H2A-Bbd_Homo          --EAQNSGERNITPLLLDMVVHNDRLLSTLFNTTTISQVAPGED----------------
H2A_T.brucei          AAAQQTKKTKRLTPRTVTLAVRHDDDLGALLRNVTMSRGGVMPSLNKALAKKQKSGKHAK
H2A.V_T.brucei        --AAKAQKTERIKPRHLLLAIRGDEELNQIV-NATIARGGVVPFVHKSLEKKIIKKSKRG
H2A_Xenopus           --RLPEITKRPVLSPGTCNSLCNDEELNKLLGGVTIAQGGVLPNIQSVLLPKKTESSKS-
H2A_Saccharomyces     --AARDNKKTRIIPRHLQLAIRNDDELNKLLGNVTIAQGGVLPNIHQNLLPKKSAKATK-
H2A_Homo              --AARDNKKTRIIPRHLQLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHK-
                                 .     . * * .    *...
```

Figure 6.1: (**A**) Crystal structure of the nucleosome core (PDB accession 1KX5) showing equivalent positions of acetylated lysine residues (red) in the C-terminal tail of histone H2A (yellow) of *T. brucei.* (**B**) Alignment of C-terminal tail regions of H2A. Acetylated *T. brucei* lysine residues are indicated with black blocks.

This degree of hyperacetylation appears to be unique to trypanosomes, as similar patterns of hyperacetylated lysine residues were also observed on the C-terminal tail of *T. cruzi* H2A (Picchi et al. 2017). Although hyperacetylation of the N-terminal tail of histone H2A variants have been observed in other eukaryotes where it was found to be associated with active transcription (Bruce 2005; Draker et al. 2012), this degree of hyperacetylation observed on the C-terminal of the core H2A histone isoform seems to be trypanosomatid specific (Picchi et al. 2017).

In *T. cruzi*, it was seen that major core histones were modified to a higher extent than their variant counterparts (Picchi et al. 2017). In *T. brucei*, the H2A.V histone variant was seen to be much more sparsely modified than H2A, with only a few modifications observed and confidently mapped (Table 3.4).

Histone H2B appear even more scarcely modified than H2A, in agreement with previous findings (Table 3.5). Only one new modification, K9me1, could be identified confidently with site determining ions on either side of the modification. The H2B.V variant was found to contain numerous combinatorial PTM patterns (Table 3.6). PTMs homologous to that observed in *T. cruzi* was seen on H2B.V. However, *T. brucei* H2B.V was modified to a greater extent than its homologous counterpart in *T. cruzi,* which was scarcely modified.

The N-terminal of histone H3 is known to be heavily modified in higher eukaryotes, exerting various epigenetic effects. However, in trypanosomes, the H3 N-terminal is hardly modified, with only a few modifications previously observed (Table 3.7), of which all were seen in

this study. In addition, tri-methylated H3K10 was detected for the first time in trypanosomes, with no mono- or di-methylation of H3K10 detected. In higher eukaryotes differential modification of the same residue can serve as a transcriptional switch, e.g. H3K9 (and H3K27) in either methylated or acetylated states can serve as transcriptional activators or repressors (Yang et al. 2003). It is not clear if trypanosomal H3K10 is the functional equivalent of H3K9 in Amorphea, where H3K9 in its tri-methylated state is associated with transcriptional repression and heterochromatin formation. Looking at the surrounding amino acid sequence context (Fig. 3.9), this does not seem to be the case. Quantitation of H3K10me3 showed that this PTM is present at low levels (approximately 1%) in both BF and PF cells. This fits with the small proportion of the genome that is transcriptionally silenced, and may indicate a possible repressive role in *T. brucei*. Unfortunately, the anti-body raised against H3K10me3 proved to be non-specific, and the genomic distribution of this modification and possible functional association with transcriptionally repressed chromatin could not be assessed. H3K11 was not observed to be modified. If *T. brucei* H3K10 is analogous to H3K9 in higher eukaryotes, tri-methylated H3K10 may function in gene repression and heterochromatin formation. Indeed, the small percentage of K10 residues observed to be tri-methylated (~1%) is in accordance with the modest section of the MBCs that needs to be transcriptionally repressed (i.e. the silent VSG arrays). The distinct amino acid sequence context surrounding the K10 residue makes it distinct from mammalian H3K9, and may present an unique and attractive epigenetic target.

H3K10me3 was found to exclusively occur in combination with H3HyPro40, forming the cPTM H3K10me3HyPro40. H3Pro40 was seen to be hydroxylated at levels of ~80% and ~60% in peptides from BF and PF life stages, respectively. Hydroxylation of H3Pro40 has interesting structural implications. Proline residues are in an equilibrium between exo and endo pucker configurations, catalysed by prolyl isomerase. In *S. cerevisiae*, H3Pro38 isomerization by FPR4 impacts methylation of H3K36 by SET2 (Nelson, Santos-Rosa, and Kouzarides 2006), as the H3Pro38 residue is in close proximity to the SET2 active site, and the ability of SET2 to structurally accommodate the exo vs. endo pucker states of H3Pro38 differs. Hydroxyproline favors the exo pucker state which prefers ϕ and ψ dihedral angles that maintains the peptide bond in a *trans* configuration (Vitagliano et al. 2001). H3HyPro40 was the only other proline residue, besides H2AHyPro26, to bear a hydroxylation modification. HILIC-MS/MS analysis of calf thymus histone spiked samples verified chemical integrity of the *T. brucei* histone samples. Absence of any hydroxylated proline residues in bovine histones spiked into samples from which *T. brucei* histones were isolated, suggested that the observed HyPro residues on *T. brucei* H2A and H3 were bona fide PTMs. Histone modification by hydroxylation has been seen in the related *T. cruzi*, where several tyrosine residues were observed to be in a hydroxylated state (Picchi et al. 2017). Although some PTMs were detected on H3.V by Mascot, only two PTMs were confidently mapped to this histone (Table 3.8). Of all the core histones, H4 and H4.V display the least sequence divergence, with 94.9% similarity and 89.9% identity shared between these isoforms (Fig. 3.5). Although some of the amino acid

substitutions can mistakenly be identified as PTMs based on observed mass shifts (e.g. D›E, V›I, and S›T as mono-methylation; A›V, K›R as di-methylation; S›E as acetylation; etc.) (Kim, Zhong, and Pandey 2016), MS analysis revealed numerous PTMs on histone H4 (Table 3.9). No combinatorial PTMs were positively identified on this histone. The H4.V isoform was also seen to be modified, but to a lesser extent than H4 (Table 3.10). Figure 6.2 shows a few of the combinatorial PTM patterns observed in this trypanosome.

MS analysis showed that *T. brucei* histones are heavily modified and possessed combinatorial PTMs of which some are analogous to PTMs seen in related kinetoplasts and model eukaryotes. These observations indicate that chromatin remodeling in trypanosomes could follow similar regulatory functions to those seen in higher eukaryotes.

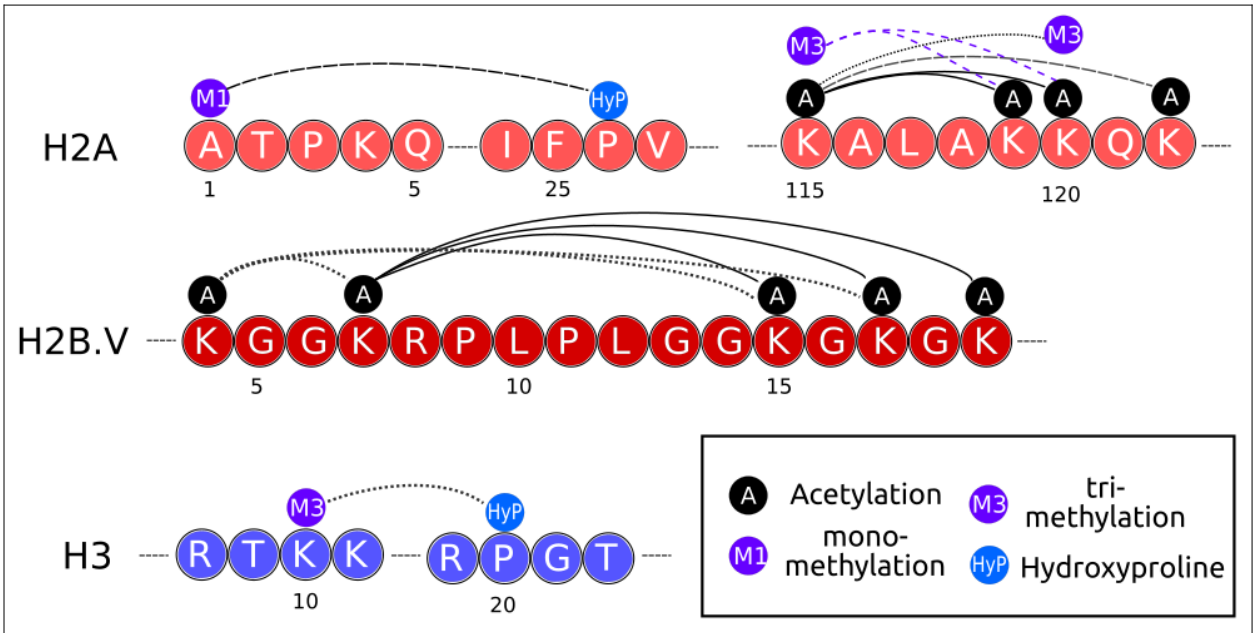


Figure 6.2: Visual representation of combinatorial PTM patterns observed on the N- and C-terminal histone tails of *T. brucei*. Distinct cPTM patterns are indicated by dashed and solid lines (e.g. H2A K115me3K119acK120ac).

Synthetic peptides were designed to raise polyclonal antibodies to map the genomic distribution of selected cPTMs by ChIP-seq and investigate possible epigenetic functions. Unfortunately, antibodies raised against tri-methylated H3K4 and K10 displayed high cross-reactivity, and α-H3HyPro40 was unable to accurately differentiate between modified and unmodified peptides. Depletion of cross-reacting antibodies by cross-adsorption did not yield any improvement.

H2AK115ac and H2AK125ac specific antibodies were able to differentiate between modified and unmodified peptides, and accurately recognized and bound to respective cPTMs. These two different cPTM patterns were chosen to investigate the genomic distribution of hyperacetylated H2A, with each antibody serving as a control for the other. This allowed interrogation of whether differentially acetylated lysine residues were used for specific epigenetic functions through formation of specific cPTM patterns, or if hyperacetylation of the H2A C-terminal tail in any combination had analogous functions.

Visual perusal of nucleosome profiles of input and ChIP DNA revealed no remarkable difference between BF and PF life stages, in agreement with previous findings where genome-wide nucleosomal architecture between life cycles was found to be remarkably similar (Maree et al. 2017). This might suggest that *T. brucei* does not extensively rely on nucleosomal architecture to regulate between life cycles. Instead, identified peaks of enrichment in the ChIP samples indicated that precise placement of nucleosomes carrying specific PTMs may be involved in epigenetic regulation of pol II transcription across life cycles.

Intersection of ChIP-seq data revealed that H2A K125ac and K115ac did indeed have different genomic distributions, indicating that hyperacetylation of the H2A C-terminal lysine residues, observed to exist in different cPTM patterns, may have different epigenetic functions. H2AK125ac was found to be enriched at SL RNA arrays (See figure 4.10), at first SASs of PTUs, and at dSSR and ISS regions - loci known for being putative pol II transcription start regions (See figure 4.11). K125ac at dSSRs extended along the entire region, and was on average 5 fold enriched over background. At ISS regions, K125ac was present downstream of the HT region, or any obstructing RNA gene that caused pol II transcriptional termination. Figure 6.3 provides an overview of H2AK125ac localization (purple) relative to pol II initiation and termination regions, as well as previously observed epigenetic marks previously that demarcate pol II PTUs.

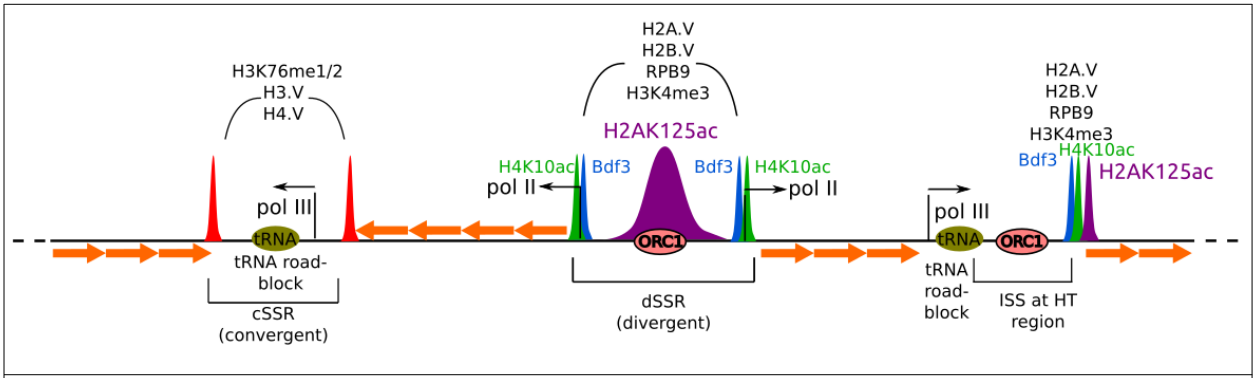


Figure 6.3: Epigenetic signals that demarcate pol II and pol III transcription units, and regulate gene expression *T. brucei*.

An assortment of epigenetic marks are known to co-localize at putative pol II transcription start sites. Comparison of H2A.V and poll II sub-unit RPB9 ChIP-seq data, both known to be enriched at pol II TSS, to loci enriched with K125ac peaks revealed a strong correlation. H2A.V and RPB9 peaks coincided with K125ac peaks at all dSSR and ISS sites investigated, observed in both BF and PF replicates (See figure 4.19). These observations suggest that hyperacetylated H2AK125ac may form part of an epigenetic mechanism responsible for recruitment of pol II transcription machinery, TSS demarcation, and chromatin compaction states. This is in line with hyperacetylation abrogating the negatively charged H2A C-terminal lysine residues which, in turn, precipitates a localized euchromatic chromatin state amenable to transcription. As K125ac peaks were observed to be enriched between RPB9 peaks, it is possible that H2A histones at dSSRs and ISSs are selectively hyperacetylated which may cause local chromatin destabilization amenable to pol II loading and bi-directional transcription initiation.

ChIP-seq data of H2AK115ac did not display any discernible distribution pattern. This cPTM was found interspersed among all the megabase chromosomes, and did not correlate with any specific gene classes or genes associated with specific life cycles. Although no evident functional association could be detected for H2AK115ac, it may perhaps have a structural rather than regulational role, serving as a binding platform for chromatin remodelers associated with nuclear architecture.

Investigation of the nuclear distribution of these cPTMs by IF microscopy revealed that H2AK115ac marks were localized to or near the nuclear

periphery (See figure 5.2), an area known to be transcriptionally repressive (Navarro, Peñate, and Landeira 2007). It might be that K115ac serves a transient role in nuclear architecture. Mini-chromosome sequences are known to localize to the nuclear periphery, as well as the H3.V histone variant which has a transcriptionally repressive effect (Lowell and Cross 2004). As K115ac displays localization foci comparable to these marks, it may be possible that this cPTM is involved in gene silencing to some extent. However, further research into the function of this cPTM is needed to unambiguously define epigenetic function. IF microscopy of K125ac revealed that this mark is distributed throughout the entire nucleus (See figure 5.1), comparable to previous work on pol II distribution (Smith et al. 2009; Daniels, Gull, and Wickstead 2010).

Numerous studies, including the current study, have shown that the epigenome is remarkably similar between different life stages in *T. brucei*. Nucleosomal architecture, histone variant deposition and histone PTMs are remarkably similar between BF and PF life stages and rarely show any life-cycle specific variation. Constitutive transcription on a genome-wide scale may be energetically inefficient but allows this parasite to rapidly respond to hostile host environments. Individual gene regulation on a genomic level requires more sophisticated mechanisms which might be temporarily ineffective in response to environmental changes, or might not have developed before the kinetoplasts diverged. It may be that genomic regulation by epigenetic mechanisms on a genome-wide scale allow faster host adaptation and contributes a certain "robustness" against environmental changes. Indeed, RNA interference

knock-down studies demonstrated that multiple layers of epigenetic mechanisms are employed in *T. brucei* to regulate genomic function, while knock-down of single epigenetic mechanisms (like H3.V, H1, HAT1, Sir2rp1) displaying only transient effects on gene expression (Alsford et al. 2007; Kawahara et al. 2008; Povelones et al. 2012; Reynolds et al. 2016). Disruption of histone PTMs does, however, result in cell-cycle defects - depletion of H3K76me2 through knock-down of the DOT1 catalyzing enzyme resulted in severe cell cycle defects like disruption of replication control, cell differentiation and antigenic variation (Janzen et al. 2006)

Overall, these observations show that *T. brucei* has a vast PTM repertoire that can be used as epigenetic tools to control genomic function. In this work, sophisticated combinatorial histone PTM patterns were observed in *T. brucei* for the first time across all histone isoforms. Corroborative findings in the related *T. cruzi* indicates that epigenetic genome regulation is an ancient process, emphasized by the intricate histone codes observed in these early branching organisms. Several of the observed (c)PTMs seem to be trypanosome specific and have been observed in the related *T. cruzi* and *L. major*, while others seem to be species specific. Some PTMs appear to be equivalent to conserved marks observed in higher eukaryotes. This hints at the establishment of epigenetic control in eukaryotic cells before divergence of kinetoplast progenitors 1.5 billion years ago. In addition, identification of epigenetic marks unique to *T. brucei* may hold tremendous promise for development of epigenetic therapies.

7) Conclusion

Over the past 20 years, mounting evidence strongly indicated that epigenetic regulation was of vital importance to trypanosomes and other related kinetoplasts (Rudenko 2010; Glover et al. 2013; Maree and Patterton 2014; De Jesus et al. 2016; Picchi et al. 2017). In *T. brucei,* a myriad of epigenetic signals, readers and writers have been shown to be crucial for proper genomic regulation, chromatin compaction, life cycle progression, host immune evasion, and virulence (Van Leeuwen et al. 1997; Mandava et al. 2007; Hughes et al. 2007; Siegel et al. 2009; Denninger et al. 2010; Povelones et al. 2012). Despite very early divergence, *T. brucei* has a sophisticated epigenetic repertoire capable of regulating a genome devoid of canonical eukaryotic transcriptional control. The number of identified *T. brucei* histone PTMs has been greatly expanded by this study: 50 individual PTMs, along with 30 distinct cPTM patterns, were observed for the first time in this trypanosome and opens up new research prospects for targeted and rational drug development. Several PTMs that were observed closely matched those seen in *T. cruzi*, some of which correlate to PTMs described in other model organisms. Other modifications appear unique to the trypanosome genus, with some PTMs unique to *T. brucei*. In addition, this study determined that *T. brucei* utilizes intricate combinatorial PTM patterns, further suggesting that a rudimentary "histone code" was present in the eukaryotic progenitors. These findings expand our knowledge of the trypanosomal (and parasitic protozoan) epigenetic repertoire. Identification of histone PTMs unique to *T. brucei*, like H3K10me3, will hopefully lead to the development of safer, targeted epigenetic therapies to treat the debilitating disease caused by *Trypanosoma brucei*.

8) References

Afgan, Enis, Dannon Baker, Bérénice Batut, Marius van den Beek, Dave Bouvier, Martin Čech, John Chilton, et al. 2018. “The Galaxy Platform for Accessible, Reproducible and Collaborative Biomedical Analyses: 2018 Update.” *Nucleic Acids Research* 46 (W1): W537–44. https://doi.org/10.1093/nar/gky379.

Afgan, Enis, Dannon Baker, Marius van den Beek, Daniel Blankenberg, Dave Bouvier, Martin Čech, John Chilton, et al. 2016. “The Galaxy Platform for Accessible, Reproducible and Collaborative Biomedical Analyses: 2016 Update.” *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkw343.

Ahn, Jeong H, Andreas Rechsteiner, Susan Strome, and William G Kelly. 2016. “A Conserved Nuclear Cyclophilin Is Required for Both RNA Polymerase II Elongation and Co-Transcriptional Splicing in Caenorhabditis Elegans.” Edited by Susan E. Mango. *PLoS Genetics* 12 (8): e1006227. https://doi.org/10.1371/journal.pgen.1006227.

Akers, Katherine G., Yoan Chérasse, Yuki Fujita, Sakthivel Srinivasan, Takeshi Sakurai, and Masanori Sakaguchi. 2018. “Concise Review: Regulatory Influence of Sleep and Epigenetics on Adult Hippocampal Neurogenesis and Cognitive and Emotional Function.” *Stem Cells*. https://doi.org/10.1002/stem.2815.

Alsford, Sam, Taemi Kawahara, Cyril Isamah, and David Horn. 2007. “A Sirtuin in the African Trypanosome Is Involved in Both DNA Repair and Telomeric Gene Silencing but Is Not Required for Antigenic Variation.” *Molecular Microbiology* 63 (3): 724–36. https://doi.org/10.1111/j.1365-2958.2006.05553.x.

Alsford, Sam, Bill Wickstead, Klaus Ersfeld, and Keith Gull. 2001. “Diversity and Dynamics of the Minichromosomal Karyotype in Trypanosoma Brucei.” *Molecular and Biochemical Parasitology*. https://doi.org/10.1016/S0166-6851(00)00388-1.

Ammar, Ron, Dax Torti, Kyle Tsui, Marinella Gebbia, Tanja Durbic, Gary D Bader, Guri Giaever, and Corey Nislow. 2012. “Chromatin Is an Ancient Innovation Conserved between Archaea and Eukarya.” *ELife* 1 (December): e00078. https://doi.org/10.7554/elife.00078.

Andrews, Simon. 2010. “FastQC: A Quality Control Tool for High Throughput Sequence Data.”

Arimura, Yasuhiro, Hiroshi Kimura, Takashi Oda, Koichi Sato, Akihisa Osakabe, Hiroaki Tachiwana, Yuko Sato, et al. 2013. “Structural Basis of a Nucleosome Containing Histone H2A.B/H2A.Bbd That Transiently Associates with Reorganized Chromatin.” *Scientific Reports* 3 (1): 3510. https://doi.org/10.1038/srep03510.

Bañuelos, Carolina P., Gabriela V. Levy, Analía G. Níttolo, Leandro G. Roser, Valeria Tekiel, and Daniel O. Sánchez. 2019. “The Trypanosoma Brucei RNA Binding Protein Tb RRM 1 Is Involved in the Transcription of a Subset of RNA Pol II -Dependent Genes.” *Journal of Eukaryotic Microbiology*, February, jeu.12716. https://doi.org/10.1111/jeu.12716.

Bao, Yunhe, Kasey Konesky, Young-Jun Park, Simona Rosu, Pamela N Dyer, Danny Rangasamy, David J Tremethick, Paul J Laybourn, and Karolin Luger. 2004. “Nucleosomes Containing the Histone Variant H2A.Bbd Organize Only 118 Base Pairs of DNA.” *The EMBO Journal* 23 (16): 3314–24. https://doi.org/10.1038/sj.emboj.7600316.

Berriman, Matthew, Elodie Ghedin, Christiane Hertz-Fowler, Gaëlle Blandin, Hubert Renauld, Daniella C Bartholomeu, Nicola J Lennard, et al. 2005. “The Genome of the African Trypanosome Trypanosoma Brucei.” *Science (New York, N.Y.)* 309 (5733): 416–22. https://doi.org/10.1126/science.1112642.

Bruce, K. 2005. “The Replacement Histone H2A.Z in a Hyperacetylated Form Is a Feature of Active Genes in the Chicken.” *Nucleic Acids Research* 33 (17): 5633–39. https://doi.org/10.1093/nar/gki874.

Cavalier-Smith, Thomas. 2010. “Kingdoms Protozoa and Chromista and the Eozoan Root of the Eukaryotic Tree.” *Biology Letters* 6 (3): 342–45. https://doi.org/10.1098/rsbl.2009.0948.

Clayton, Christine. 2019. “Regulation of Gene Expression in Trypanosomatids: Living with Polycistronic Transcription.” *Open Biology* 9 (6): 190072. https://doi.org/10.1098/rsob.190072.

Cole, Hope A., Bruce H. Howard, and David J. Clark. 2012. “Genome-Wide Mapping of Nucleosomes in Yeast Using Paired-End Sequencing.” In , 145–68. https://doi.org/10.1016/B978-0-12-391938-0.00006-9.

Crowe, J.S., A.G. Lamont, J.D. Barry, and K. Vickerman. 1984. “Cytotoxicity of Monoclonal Antibodies to Trypanosoma Brucei.” *Transactions of the Royal Society of Tropical Medicine and Hygiene* 78 (4): 508–13. https://doi.org/10.1016/0035-9203(84)90073-7.

Cui, Liwang, and Jun Miao. 2010. “Chromatin-Mediated Epigenetic Regulation in the Malaria Parasite Plasmodium Falciparum.” *Eukaryotic Cell* 9 (8): 1138–49. https://doi.org/10.1128/EC.00036-10.

Daniels, J.-P., Keith Gull, and Bill Wickstead. 2010. “Cell Biology of the Trypanosome Genome.” *Microbiology and Molecular Biology Reviews* 74 (4): 552–69. https://doi.org/10.1128/MMBR.00024-10.

Dechat, T., S. A. Adam, P. Taimen, T. Shimi, and R. D. Goldman. 2010. “Nuclear Lamins.” *Cold Spring Harbor Perspectives in Biology* 2 (11): a000547–a000547. https://doi.org/10.1101/cshperspect.a000547.

Denninger, Viola, Alexander Fullbrook, Mohamed Bessat, Klaus Ersfeld, and Gloria Rudenko. 2010. “The FACT Subunit TbSpt16 Is Involved in Cell Cycle Specific Control of VSG Expression Sites in Trypanosoma

Brucei." *Molecular Microbiology* 78 (2): 459–74. https://doi.org/10.1111/j.1365-2958.2010.07350.x.

Draker, Ryan, Marlee K. Ng, Elizabeth Sarcinella, Vladimir Ignatchenko, Thomas Kislinger, and Peter Cheung. 2012. "A Combination of H2A.Z and H4 Acetylation Recruits Brd2 to Chromatin during Transcriptional Activation." Edited by Jerry Workman. *PLoS Genetics* 8 (11): e1003047. https://doi.org/10.1371/journal.pgen.1003047.

Droll, Dorothea, Igor Minia, Abeer Fadda, Aditi Singh, Mhairi Stewart, Rafael Queiroz, and Christine Clayton. 2013. "Post-Transcriptional Regulation of the Trypanosome Heat Shock Response by a Zinc Finger Protein." *PLoS Pathogens* 9 (4). https://doi.org/10.1371/journal.ppat.1003286.

Echeverry, Maria C, Christopher Bot, Samson O Obado, Martin C Taylor, and John M Kelly. 2012. "Centromere-Associated Repeat Arrays on Trypanosoma Brucei Chromosomes Are Much More Extensive than Predicted." *BMC Genomics* 13 (1): 29. https://doi.org/10.1186/1471-2164-13-29.

Ersfeld, Klaus. 2011. "Nuclear Architecture, Genome and Chromatin Organisation in Trypanosoma Brucei." *Research in Microbiology* 162 (6): 626–36. https://doi.org/10.1016/j.resmic.2011.01.014.

Fèvre, Eric M., Beatrix V. Wissmann, Susan C Welburn, and Pascal Lutumba. 2008. "The Burden of Human African Trypanosomiasis." Edited by Simon Brooker. *PLoS Neglected Tropical Diseases* 2 (12): e333. https://doi.org/10.1371/journal.pntd.0000333.

Figueiredo, Luisa M., Christian J. Janzen, and George A M Cross. 2008. "A Histone Methyltransferase Modulates Antigenic Variation in African Trypanosomes." *PLoS Biology*. https://doi.org/10.1371/journal.pbio.0060161.

Figueiredo, Luisa M, and George A M Cross. 2010. “Nucleosomes Are Depleted at the VSG Expression Site Transcribed by RNA Polymerase I in African Trypanosomes.” *Eukaryotic Cell* 9 (1): 148–54. https://doi.org/10.1128/EC.00282-09.

Figueiredo, Luisa M, George A M Cross, and Christian J Janzen. 2009. “Epigenetic Regulation in African Trypanosomes: A New Kid on the Block.” *Nature Reviews. Microbiology* 7 (7): 504–13. https://doi.org/10.1038/nrmicro2149.

Finlan, Lee E., Duncan Sproul, Inga Thomson, Shelagh Boyle, Elizabeth Kerr, Paul Perry, Bauke Ylstra, Jonathan R. Chubb, and Wendy A. Bickmore. 2008. “Recruitment to the Nuclear Periphery Can Alter Expression of Genes in Human Cells.” Edited by Wolf Reik. *PLoS Genetics* 4 (3): e1000039. https://doi.org/10.1371/journal.pgen.1000039.

García-Salcedo, José A., Purificación Gijón, Derek P. Nolan, Patricia Tebabi, and Etienne Pays. 2003. “A Chromosomal SIR2 Homologue with Both Histone NAD-Dependent ADP-Ribosyltransferase and Deacetylase Activities Is Involved in DNA Repair in Trypanosoma Brucei.” *EMBO Journal* 22 (21): 5851–62. https://doi.org/10.1093/emboj/cdg553.

Garcia, Benjamin A., Sandra B. Hake, Robert L. Diaz, Monika Kauer, Stephanie A. Morris, Judith Recht, Jeffrey Shabanowitz, et al. 2007. “Organismal Differences in Post-Translational Modifications in Histones H3 and H4.” *Journal of Biological Chemistry* 282 (10): 7641–55. https://doi.org/10.1074/jbc.M607900200.

Gassen, Alwine, Doris Brechtefeld, Niklas Schandry, J. Manuel Arteaga-Salas, Lars Israel, Axel Imhof, and Christian J. Janzen. 2012. “DOT1A-Dependent H3K76 Methylation Is Required for Replication Regulation in Trypanosoma Brucei.” *Nucleic Acids Research* 40 (20): 10302–11. https://doi.org/10.1093/nar/gks801.

Gasteiger, Elisabeth, Christine Hoogland, Alexandre Gattiker, S'everine Duvaud, Marc R. Wilkins, Ron D. Appel, and Amos Bairoch. 2005. "Protein Identification and Analysis Tools on the ExPASy Server." In *The Proteomics Protocols Handbook*, 571–607. Totowa, NJ: Humana Press. https://doi.org/10.1385/1-59259-890-0:571.

Glover, Lucy, Sebastian Hutchinson, Sam Alsford, Richard Mcculloch, Mark C. Field, and David Horn. 2013. "Antigenic Variation in African Trypanosomes: The Importance of Chromosomal and Nuclear Context in VSG Expression Control." *Cellular Microbiology* 15 (12): 1984–93. https://doi.org/10.1111/cmi.12215.

Gorres, Kelly L., and Ronald T. Raines. 2010. "Prolyl 4-Hydroxylase." *Critical Reviews in Biochemistry and Molecular Biology* 45 (2): 106–24. https://doi.org/10.3109/10409231003627991.

Guillemette, Benoit, Paul Drogaris, Hsiu-Hsu Sophia Lin, Harry Armstrong, Kyoko Hiragami-Hamada, Axel Imhof, Éric Bonneil, Pierre Thibault, Alain Verreault, and Richard J. Festenstein. 2011. "H3 Lysine 4 Is Acetylated at Active Gene Promoters and Is Regulated by H3 Lysine 4 Methylation." Edited by Hiten D. Madhani. *PLoS Genetics* 7 (3): e1001354. https://doi.org/10.1371/journal.pgen.1001354.

Günzl, Arthur, Thomas Bruderer, Gabriele Laufer, Bernd Schimanski, Lan-chun Tu, Hui-min Chung, Pei-Tseng Lee, and Mary Gwo-shu Lee. 2003. "RNA Polymerase I Transcribes Procyclin Genes and Variant Surface Glycoprotein Gene Expression Sites in Trypanosoma Brucei." *Eukaryotic Cell* 2 (3): 542–51. https://doi.org/10.1128/EC.2.3.542.

Hake, Sandra B, Benjamin A Garcia, Elizabeth M Duncan, Monika Kauer, Graham Dellaire, Jeffrey Shabanowitz, David P. Bazett-Jones, C David Allis, and Donald F Hunt. 2006. "Expression Patterns and Post-Translational Modifications Associated with Mammalian Histone H3 Variants." *Journal of Biological Chemistry* 281 (1): 559–68. https://doi.org/10.1074/jbc.M509266200.

He, Ding, Omar Fiz-Palacios, Cheng Jie Fu, Chun Chieh Tsai, and Sandra L. Baldauf. 2014. “An Alternative Root for the Eukaryote Tree of Life.” *Current Biology* 24 (4): 465–70. https://doi.org/10.1016/j.cub.2014.01.036.

Herbert, W. J., and W. H R Lumsden. 1976. “Trypanosoma Brucei: A Rapid ‘Matching’ Method for Estimating the Host’s Parasitemia.” *Experimental Parasitology*. https://doi.org/10.1016/0014-4894(76)90110-7.

Hirumi, H, and Kazuko Hirumi. 1989. “Continuous Cultivation of Trypanosoma Brucei Blood Stream Forms in a Medium Containing a Low Concentration of Serum Protein without Feeder Cell Layers.” *The Journal of Parasitology* 75 (6): 985–89. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=2614608.

Huber, Michael D., and Larry Gerace. 2007. “The Size-Wise Nucleus: Nuclear Volume Control in Eukaryotes.” *Journal of Cell Biology*. https://doi.org/10.1083/jcb.200710156.

Hughes, Katie, Matthew Wand, Lucy Foulston, Rosanna Young, Kate Harley, Stephen Terry, Klaus Ersfeld, and Gloria Rudenko. 2007. “A Novel ISWI Is Involved in VSG Expression Site Downregulation in African Trypanosomes.” *The EMBO Journal* 26 (9): 2400–2410. https://doi.org/10.1038/sj.emboj.7601678.

Ismail, Hamid D., Robert H. Newman, and Dukka B. KC. 2016. “RF-Hydroxysite: A Random Forest Based Predictor for Hydroxylation Sites.” *Molecular BioSystems* 12 (8): 2427–35. https://doi.org/10.1039/C6MB00179C.

Janzen, Christian J, Joseph P Fernandez, Haiteng Deng, Robert Diaz, Sandra B Hake, and George A M Cross. 2006. “Unusual Histone Modifications in Trypanosoma Brucei.” *FEBS Letters* 580 (9): 2306–10. https://doi.org/10.1016/j.febslet.2006.03.044.

Jenkins, Cara L., Lynn E. Bretscher, Ilia A. Guzei, and Ronald T. Raines. 2003. “Effect of 3-Hydroxyproline Residues on Collagen Stability.” *Journal of the American Chemical Society* 125 (21): 6422–27. https://doi.org/10.1021/ja034015j.

Jesus, Teresa Cristina Leandro De, Vinícius Santana Nunes, Mariana De Camargo Lopes, Daiana Evelin Martil, Leo Kei Iwai, Nilmar Silvio Moretti, Fabrício Castro MacHado, et al. 2016. “Chromatin Proteomics Reveals Variable Histone Modifications during the Life Cycle of Trypanosoma Cruzi.” *Journal of Proteome Research* 15 (6): 2039–51. https://doi.org/10.1021/acs.jproteome.6b00208.

Jha, Pravin K., Mohd. Imran Khan, Anshul Mishra, Pradeep Das, and Kislay K. Sinha. 2017. “HAT2 Mediates Histone H4K4 Acetylation and Affects Micrococcal Nuclease Sensitivity of Chromatin in Leishmania Donovani.” Edited by Axel Imhof. *PLOS ONE* 12 (5): e0177372. https://doi.org/10.1371/journal.pone.0177372.

Kawahara, Taemi, T. Nicolai Siegel, Alexandra K. Ingram, Sam Alsford, G. A M Cross, and David Horn. 2008. “Two Essential MYST-Family Proteins Display Distinct Roles in Histone H4K10 Acetylation and Telomeric Silencing in Trypanosomes.” *Molecular Microbiology*. https://doi.org/10.1111/j.1365-2958.2008.06346.x.

Kim, Min-Sik, Jun Zhong, and Akhilesh Pandey. 2016. “Common Errors in Mass Spectrometry-Based Analysis of Post-Translational Modifications.” *PROTEOMICS* 16 (5): 700–714. https://doi.org/10.1002/pmic.201500355.

Knoll, A. H., E. J. Javaux, D. Hewitt, and P. Cohen. 2006. “Eukaryotic Organisms in Proterozoic Oceans.” *Philosophical Transactions of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rstb.2006.1843.

Kossel A. 1884. “Uber Einen Peptonartigen Bestandteil Des Zellkerns.” *Z Physiol. Chem* 8: 511–15.

https://www.degruyter.com/view/j/bchm1.1884.8.issue-6/bchm1.1884.8.6.511/bchm1.1884.8.6.511.xml.

Kouzarides, Tony. 2007. “Chromatin Modifications and Their Function.” *Cell* 128 (4): 693–705. https://doi.org/10.1016/j.cell.2007.02.005.

Lachner, Monika, and Thomas Jenuwein. 2002. “The Many Faces of Histone Lysine Methylation.” *Current Opinion in Cell Biology* 14 (3): 286–98. https://www.sciencedirect.com/science/article/pii/S0955067402003356.

Langmead, Ben, Cole Trapnell, Mihai Pop, and SL Salzberg. 2009. “Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome.” *Genome Biol.* 10 (3): R25. https://doi.org/10.1186/gb-2009-10-3-r25.

Lanham, Sheila M., and D. G. Godfrey. 1970. “Isolation of Salivarian Trypanosomes from Man and Other Mammals Using DEAE-Cellulose.” *Experimental Parasitology*. https://doi.org/10.1016/0014-4894(70)90120-7.

Leeuwen, Fred Van, Eric R. Wijsman, Rudo Kieft, Gijs A. Van Der Marel, Jacques H. Van Boom, and Piet Borst. 1997. “Localization of the Modified Base J in Telomeric VSG Gene Expression Sites of Trypanosoma Brucei.” *Genes and Development* 11 (23): 3232–41. https://doi.org/10.1101/gad.11.23.3232.

Lowell, Joanna E, and George A M Cross. 2004. “A Variant Histone H3 Is Enriched at Telomeres in Trypanosoma Brucei.” *Journal of Cell Science* 117 (Pt 24): 5937–47. https://doi.org/10.1242/jcs.01515.

Luger, Karolin, Robin K Richmond, David F Sargent, Timothy J Richmond, and Armin W A. 1997. “Crystal Structure of the Nucleosome Core Particle at 2 . 8 A Resolution.” *Nature* 389: 251–60. https://doi.org/10.1038/38444.

Magklara, Angeliki, Angela Yen, Bradley M. Colquitt, E. Josephine Clowney, William Allen, Eirene Markenscoff-Papadimitriou, Zoe A. Evans, et al. 2011. “An Epigenetic Signature for Monoallelic Olfactory Receptor Expression.” *Cell* 145 (4): 555–70. https://doi.org/10.1016/j.cell.2011.03.040.

Mandava, Veena, Joseph P. Fernandez, Haiteng Deng, Christian J. Janzen, Sandra B. Hake, and George A.M. Cross. 2007. “Histone Modifications in Trypanosoma Brucei.” *Molecular and Biochemical Parasitology* 156 (1): 41–50. https://doi.org/10.1016/j.molbiopara.2007.07.005.

Mandava, Veena, Christian J Janzen, and George A M Cross. 2008. “Trypanosome H2Bv Replaces H2B in Nucleosomes Enriched for H3 K4 and K76 Trimethylation.” *Biochemical and Biophysical Research Communications* 368 (4): 846–51. https://doi.org/10.1016/j.bbrc.2008.01.144.

Maree, Johannes P., and Hugh G. Patterton. 2014. “The Epigenome of Trypanosoma Brucei: A Regulatory Interface to an Unconventional Transcriptional Machine.” *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 1839 (9): 743–50. https://doi.org/10.1016/j.bbagrm.2014.05.028.

Maree, Johannes Petrus, Megan Lindsay Povelones, David Johannes Clark, Gloria Rudenko, and Hugh-George Patterton. 2017. “Well-Positioned Nucleosomes Punctuate Polycistronic Pol II Transcription Units and Flank Silent VSG Gene Arrays in Trypanosoma Brucei.” *Epigenetics & Chromatin* 10 (1): 14. https://doi.org/10.1186/s13072-017-0121-9.

Martínez-Calvillo, Santiago, Juan C. Vizuet-De-Rueda, Luis E. Florencio-Martínez, Rebeca G. Manning-Cela, and Elisa E. Figueroa-Angulo. 2010. “Gene Expression in Trypanosomatid Parasites.” *Journal of Biomedicine and Biotechnology*. https://doi.org/10.1155/2010/525241.

McAndrew, M, Shiela Graham, Claudia Hartmann, and Christine Clayton. 1998. “Testing Promoter Activity in the Trypanosome Genome: Isolation of a Metacyclic-Type VSG Promoter, and Unexpected Insights into RNA Polymerase II Transcription.” *Experimental Parasitology* 90 (1): 65–76. https://doi.org/10.1006/expr.1998.4317.

Mekhail, Karim, and Danesh Moazed. 2010. “The Nuclear Envelope in Genome Organization, Expression and Stability.” *Nature Reviews Molecular Cell Biology* 11 (5): 317–28. https://doi.org/10.1038/nrm2894.

Moretti, Nilmar Silvio, Igor Cestari, Atashi Anupama, Ken Stuart, and Sergio Schenkman. 2018. “Comparative Proteomic Analysis of Lysine Acetylation in Trypanosomes.” *Journal of Proteome Research* 17 (1): 374–85. https://doi.org/10.1021/acs.jproteome.7b00603.

Myllyharju, Johanna. 2003. “Prolyl 4-Hydroxylases, the Key Enzymes of Collagen Biosynthesis.” *Matrix Biology* 22 (1): 15–24. https://doi.org/10.1016/S0945-053X(03)00006-4.

Narayanan, Mani Shankar, and Gloria Rudenko. 2013. “TDP1 Is an HMG Chromatin Protein Facilitating RNA Polymerase I Transcription in African Trypanosomes.” *Nucleic Acids Research* 41 (5): 2981–92. https://doi.org/10.1093/nar/gks1469.

Nardelli, Sheila C., Li-Min Ting, and Kami Kim. 2015. “Techniques to Study Epigenetic Control and the Epigenome in Parasites.” In *Springer*, 177–91. https://doi.org/10.1007/978-1-4939-1438-8_10.

Nardelli, Sheila Cristina, Julia Pinheiro Chagas da Cunha, Maria Cristina M. Motta, and Sergio Schenkman. 2009. “Distinct Acetylation of Trypanosoma Cruzi Histone H4 during Cell Cycle, Parasite Differentiation, and after DNA Damage.” *Chromosoma* 118 (4): 487–99. https://doi.org/10.1007/s00412-009-0213-9.

Navarro, Miguel, and Keith Gull. 2001. “A Pol I Transcriptional Body Associated with VSG Mono-Allelic Expression in Trypanosoma Brucei.” *Nature* 414 (6865): 759–63. https://doi.org/10.1038/414759a.

Navarro, Miguel, Xenia Peñate, and David Landeira. 2007. “Nuclear Architecture Underlying Gene Expression in Trypanosoma Brucei.” *Trends in Microbiology* 15 (6): 263–70. https://doi.org/10.1016/j.tim.2007.04.004.

Neelin, J. M., P. X. Callahan, D. C. Lamb, and K. Murray. 1964. “THE HISTONES OF CHICKEN ERYTHROCYTE NUCLEI.” *Canadian Journal of Biochemistry* 42 (12): 1743–52. https://doi.org/10.1139/o64-185.

Nelson, Christopher J., Helena Santos-Rosa, and Tony Kouzarides. 2006. “Proline Isomerization of Histone H3 Regulates Lysine Methylation and Gene Expression.” *Cell* 126 (5): 905–16. https://doi.org/10.1016/j.cell.2006.07.026.

Perdivara, Irina, Leesa J. Deterding, Michael Przybylski, and Kenneth B. Tomer. 2010. “Mass Spectrometric Identification of Oxidative Modifications of Tryptophan Residues in Proteins: Chemical Artifact or Post-Translational Modification?” *Journal of the American Society for Mass Spectrometry* 21 (7): 1114–17. https://doi.org/10.1016/j.jasms.2010.02.016.

Pesavento, James J., Craig A. Mizzen, and Neil L. Kelleher. 2006. “Quantitative Analysis of Modified Proteins and Their Positional Isomers by Tandem Mass Spectrometry: Human Histone H4.” *Analytical Chemistry* 78 (13): 4271–80. https://doi.org/10.1021/ac0600050.

Picchi, Gisele F.A., Vanessa Zulkievicz, Marco A. Krieger, Nilson T. Zanchin, Samuel Goldenberg, and Lyris M.F. De Godoy. 2017. “Post-Translational Modifications of Trypanosoma Cruzi Canonical and Variant Histones.” *Journal of Proteome Research* 16 (3): 1167–79. https://doi.org/10.1021/acs.jproteome.6b00655.

Poon, S. K., L. Peacock, W. Gibson, K. Gull, and S. Kelly. 2012. “A Modular and Optimized Single Marker System for Generating Trypanosoma Brucei Cell Lines Expressing T7 RNA Polymerase and the Tetracycline Repressor.” *Open Biology* 2 (2): 110037–110037. https://doi.org/10.1098/rsob.110037.

Povelones, Megan L, Eva Gluenz, Marcin Dembek, Keith Gull, and Gloria Rudenko. 2012. “Histone H1 Plays a Role in Heterochromatin Formation and VSG Expression Site Silencing in Trypanosoma Brucei.” Edited by Elisabetta Ullu. *PLoS Pathogens* 8 (11): e1003010. https://doi.org/10.1371/journal.ppat.1003010.

Ramos, Thiago Cesar Prata, Vinicius Santana Nunes, Sheila Cristina Nardelli, Bruno dos Santos Pascoalino, Nilmar Silvio Moretti, Antonio Augusto Rocha, Leonardo da Silva Augusto, and Sergio Schenkman. 2015. “Expression of Non-Acetylatable Lysines 10 and 14 of Histone H4 Impairs Transcription and Replication in Trypanosoma Cruzi.” *Molecular and Biochemical Parasitology* 204 (1): 1–10. https://doi.org/10.1016/j.molbiopara.2015.11.001.

Reynolds, David, Brigitte T Hofmeister, Laura Cliffe, Magdy Alabady, T Nicolai Siegel, Robert J Schmitz, and Robert Sabatini. 2016. “Histone H3 Variant Regulates RNA Polymerase II Transcription Termination and Dual Strand Transcription of SiRNA Loci in Trypanosoma Brucei.” Edited by Luisa Figueiredo. *PLoS Genetics* 12 (1): e1005758. https://doi.org/10.1371/journal.pgen.1005758.

Robinson, James T, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S Lander, Gad Getz, and Jill P Mesirov. 2011. “Integrative Genomics Viewer.” *Nature Biotechnology*. https://doi.org/10.1038/nbt.1754.

Rudenko, Gloria. 2010. “Epigenetics and Transcriptional Control in African Trypanosomes.” *Essays in Biochemistry* 48 (1): 201–19. https://doi.org/10.1017/CBO9781107415324.004.

Schneider, Caroline A, Wayne S Rasband, and Kevin W Eliceiri. 2012. “NIH Image to ImageJ: 25 Years of Image Analysis.” *Nature Methods*.

Shechter, D, H L Dormann, C D Allis, and S B Hake. 2007. “Extraction, Purification and Analysis of Histones: Abstract: Nature Protocols.” *Nature Protocols*. https://www.nature.com/nprot/journal/v2/n6/abs/nprot.2007.202.html.

Sidoli, Simone, Lei Cheng, and Ole N. Jensen. 2012. “Proteomics in Chromatin Biology and Epigenetics: Elucidation of Post-Translational Modifications of Histone Proteins by Mass Spectrometry.” *Journal of Proteomics* 75 (12): 3419–33. https://doi.org/10.1016/j.jprot.2011.12.029.

Sidoli, Simone, Veit Schwämmle, Chrystian Ruminowicz, Thomas A. Hansen, Xudong Wu, Kristian Helin, and Ole N. Jensen. 2014. “Middle-down Hybrid Chromatography/Tandem Mass Spectrometry Workflow for Characterization of Combinatorial Post-Translational Modifications in Histones.” *PROTEOMICS* 14 (19): 2200–2211. https://doi.org/10.1002/pmic.201400084.

Siegel, T. Nicolai, Taemi Kawahara, Jeffrey A. DeGrasse, Christian J. Janzen, David Horn, and G. A M Cross. 2008. “Acetylation of Histone H4K4 Is Cell Cycle Regulated and Mediated by HAT3 in Trypanosoma Brucei.” *Molecular Microbiology* 67 (4): 762–71. https://doi.org/10.1111/j.1365-2958.2007.06079.x.

Siegel, T Nicolai, Doeke R Hekstra, Louise E Kemp, Luisa M Figueiredo, Joanna E Lowell, David Fenyo, Xuning Wang, Scott Dewell, and George A M Cross. 2009. “Four Histone Variants Mark the Boundaries of Polycistronic Transcription Units in Trypanosoma Brucei.” *Genes & Development* 23 (9): 1063–76. https://doi.org/10.1101/gad.1790409.

Sievers, Fabian, Andreas Wilm, David Dineen, Toby J. Gibson, Kevin Karplus, Weizhong Li, Rodrigo Lopez, et al. 2011. “Fast, Scalable

Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega." *Molecular Systems Biology*. https://doi.org/10.1038/msb.2011.75.

Smith, Terry K., Nadina Vasileva, Eva Gluenz, Stephen Terry, Neil Portman, Susanne Kramer, Mark Carrington, Shulamit Michaeli, Keith Gull, and Gloria Rudenko. 2009. "Blocking Variant Surface Glycoprotein Synthesis in Trypanosoma Brucei Triggers a General Arrest in Translation Initiation." Edited by Nina Papavasiliou. *PLoS ONE* 4 (10): e7532. https://doi.org/10.1371/journal.pone.0007532.

Strahl, B D, and C D Allis. 2000. "The Language of Covalent Histone Modifications." *Nature* 403 (6765): 41–45. https://doi.org/10.1038/47412.

Su, Zhangli, Melissa D Boersma, Jin-Hee Lee, Samuel S Oliver, Shichong Liu, Benjamin A Garcia, and John M Denu. 2014. "ChIP-Less Analysis of Chromatin States." *Epigenetics & Chromatin* 7: 7. https://doi.org/10.1186/1756-8935-7-7.

Su, Zhangli, and John M. Denu. 2016. "Reading the Combinatorial Histone Language." *ACS Chemical Biology* 11 (3): 564–74. https://doi.org/10.1021/acschembio.5b00864.

Talbert, Paul B., and Steven Henikoff. 2009. "Chromatin-Based Transcriptional Punctuation." *Genes and Development* 23 (9): 1037–41. https://doi.org/10.1101/gad.1806409.

Taylor, Jesse E., and Gloria Rudenko. 2006. "Switching Trypanosome Coats: What's in the Wardrobe?" *Trends in Genetics* 22 (11): 614–20. https://doi.org/10.1016/j.tig.2006.08.003.

Thorvaldsdóttir, Helga, James T. Robinson, and Jill P. Mesirov. 2013. "Integrative Genomics Viewer (IGV): High-Performance Genomics Data Visualization and Exploration." *Briefings in Bioinformatics* 14 (2): 178–92. https://doi.org/10.1093/bib/bbs017.

Vargas-Parada, L. 2010. “Kinetoplastids and Their Networks of Interlocked DNA.” *Nature Education*.

Verrastro, Ivan, Sabah Pasha, Karina Jensen, Andrew Pitt, and Corinne Spickett. 2015. “Mass Spectrometry-Based Methods for Identifying Oxidized Proteins in Disease: Advances and Challenges.” *Biomolecules* 5 (2): 378–411. https://doi.org/10.3390/biom5020378.

Vitagliano, Luigi, Rita Berisio, Lelio Mazzarella, and Adriana Zagari. 2001. “Structural Bases of Collagen Stabilization Induced by Proline Hydroxylation.” *Biopolymers*. https://doi.org/10.1002/1097-0282(20010415)58:5<459::aid-bip1021>3.0.co;2-v.

Wedel, Carolin, Konrad U Förstner, Ramona Derr, and T Nicolai Siegel. 2017. “GT-rich Promoters Can Drive RNA Pol II Transcription and Deposition of H2A.Z in African Trypanosomes.” *The EMBO Journal* 36 (17): 2581–94. https://doi.org/10.15252/embj.201695323.

Wel, Hanke van der, Howard R. Morris, Maria Panico, Thanai Paxton, Anne Dell, Lee Kaplan, and Christopher M. West. 2002. “Molecular Cloning and Expression of a UDP- N -Acetylglucosamine (GlcNAc):Hydroxyproline Polypeptide GlcNAc-Transferase That Modifies Skp1 in the Cytoplasm of Dictyostelium.” *Journal of Biological Chemistry* 277 (48): 46328–37. https://doi.org/10.1074/jbc.M208024200.

Wente, S. R., and M. P. Rout. 2010. “The Nuclear Pore Complex and Nuclear Transport.” *Cold Spring Harbor Perspectives in Biology* 2 (10): a000562–a000562. https://doi.org/10.1101/cshperspect.a000562.

West, Christopher M., and Ira J. Blader. 2015. “Oxygen Sensing by Protozoans: How They Catch Their Breath.” *Current Opinion in Microbiology* 26: 41–47. https://doi.org/10.1016/j.mib.2015.04.006.

Wheeler, Richard. n.d. “Lifecycle of T. Brucei as It Progresses from Fly to Man and Back Again.” wheelerlab.net.

WHO. 2015. “Trypanosomiasis , Human African (Sleeping Sickness).” Who Factsheets. 2015. https://www.who.int/news-room/fact-sheets/detail/trypanosomiasis-human-african-(sleeping-sickness).

Wickstead, Bill, Klaus Ersfeld, and Keith Gull. 2004. “The Small Chromosomes of Trypanosoma Brucei Involved in Antigenic Variation Are Constructed around Repetitive Palindromes.” *Genome Research* 14 (6): 1014–24. https://doi.org/10.1101/gr.2227704.

Wolffe, A. 1998. “Chromatin, San Diego.” In .

World Health Organization. 2013. “Control and Surveillance of Human African Trypanosomiasis.” *World Health Organization Technical Report Series*, no. 984: 1–237. http://www.ncbi.nlm.nih.gov/pubmed/24552089.

Wright, Jessica R., T. Nicolai Siegel, and George A M Cross. 2010. “Histone H3 Trimethylated at Lysine 4 Is Enriched at Probable Transcription Start Sites in Trypanosoma Brucei.” *Molecular and Biochemical Parasitology* 172 (2): 141–44. https://doi.org/10.1016/j.molbiopara.2010.03.013.

Wyse, Brandon A., Roxanne Oshidari, Daniel C.B. Jeffery, and Krassimir Y. Yankulov. 2013. “Parasite Epigenetics and Immune Evasion: Lessons from Budding Yeast.” *Epigenetics and Chromatin* 6 (1): 40. https://doi.org/10.1186/1756-8935-6-40.

Yang, Xiao, Xuelei Wu, Jiahai Zhang, Xuecheng Zhang, Chao Xu, Shanhui Liao, and Xiaoming Tu. 2017. “Recognition of Hyperacetylated N-Terminus of H2AZ by TbBDF2 from Trypanosoma Brucei.” *Biochemical Journal* 474 (22): 3817–30. https://doi.org/10.1042/BCJ20170619.

Yang, Youwen, Tao Li, Thanh H. Vu, Gary A. Ulaner, Ji Fan Hu, and Andrew R. Hoffman. 2003. “The Histone Code Regulating Expression of the

Imprinted Mouse Igf2r Gene." *Endocrinology*. https://doi.org/10.1210/en.2003-0798.

Young, Nicolas L, Peter A. DiMaggio, Mariana D. Plazas-Mayorca, Richard C Baliban, Christodoulos A Floudas, and Benjamin A. Garcia. 2009. "High Throughput Characterization of Combinatorial Histone Codes." *Molecular & Cellular Proteomics* 8 (10): 2266–84. https://doi.org/10.1074/mcp.M900238-MCP200.

Zhang, Yong, Tao Liu, Clifford A Meyer, Jérôme Eeckhoute, David S Johnson, Bradley E Bernstein, Chad Nusbaum, et al. 2008. "Model-Based Analysis of ChIP-Seq (MACS)." *Genome Biology*. https://doi.org/10.1186/gb-2008-9-9-r137.

Zhao, Yingming, and Benjamin A. Garcia. 2015. "Comprehensive Catalog of Currently Documented Histone Modifications." *Cold Spring Harbor Perspectives in Biology* 7 (9): a025064. https://doi.org/10.1101/cshperspect.a025064.

9) Addenda

Supplementary document R1: *In silico* digestions of *T. brucei* core and variant histones by Glu-C or Asp-N endo-proteinases.

Histone H2A:

Name of enzyme	No. of cleavages	Positions of cleavage sites
Glu-C	3	56, 61, 64

```
                                                                Glu
                                                                 |
      ATPKQAVKKASKGGSSRSVKAGLIFPVGRVGTLLRRGQYARRIGASGAVYMAAVLEYLTA
   1  ---------+---------+---------+---------+---------+---------+   60

        Glu
    Glu  |
      |  |
      ELLELSVKAAAQQTKKTKRLTPRTVTLAVRHDDDLGALLRNVTMSRGGVMPSLNKALAKK
  61  ---------+---------+---------+---------+---------+---------+   120


      QKSGKHAKATPSV
 121  ---------+---   133
```

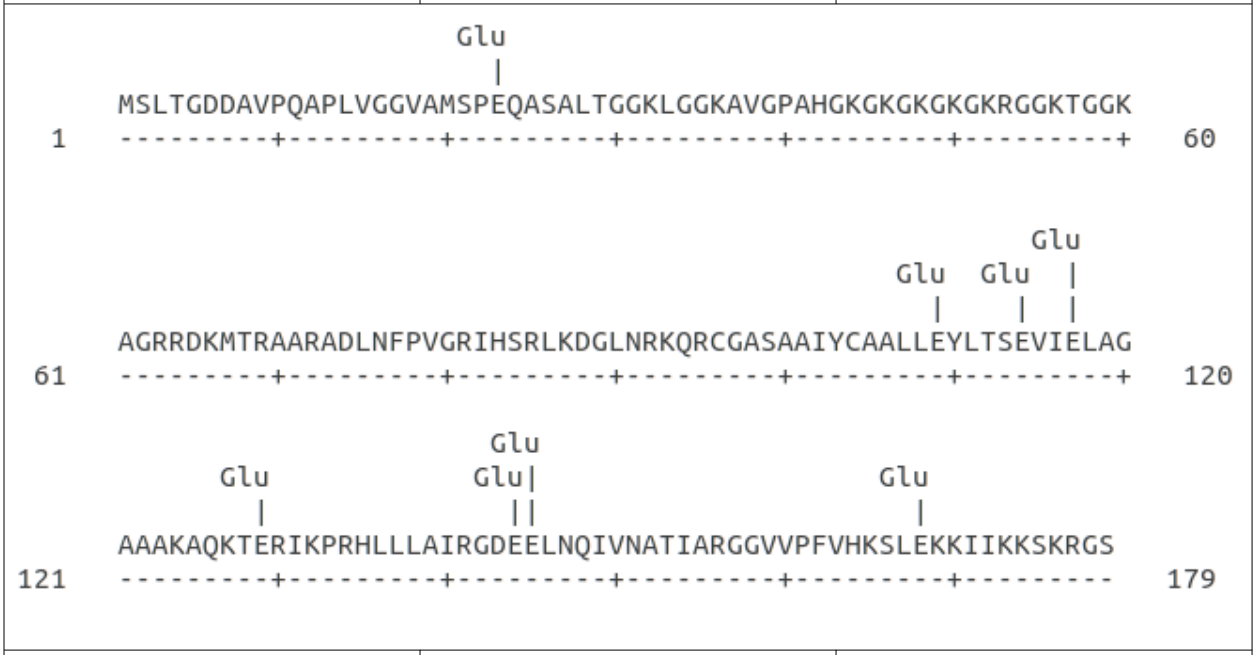
Name of enzyme	No. of cleavages	Positions of cleavage sites
Asp-N	3	92, 93, 94

```
      MATPKQAVKKASKGGSSRSVKAGLIFPVGRVGTLLRRGQYARRIGASGAVYMAAVLEYLT
   1  ---------+---------+---------+---------+---------+---------+   60

                                AspN
                               AspN|
                              AspN||
                                |||
      AELLELSVKAAAQQTKKTKRLTPRTVTLAVRHDDDLGALLRNVTMSRGGVMPSLNKALAK
  61  ---------+---------+---------+---------+---------+---------+   120


      KQKSGKHAKATPSV
 121  ---------+----   134
```

Supplementary document R1 (continued):

Histone H2A.V:

Name of enzyme	No. of cleavages	Positions of cleavage sites
Glu-C	8	23, 109, 114, 117, 129, 144, 145, 168

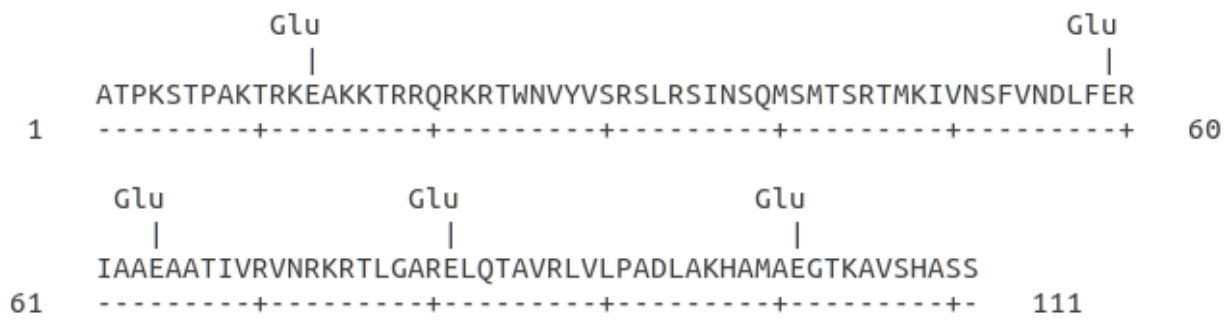


Name of enzyme	No. of cleavages	Positions of cleavage sites
Asp-N	6	5, 6, 64, 73, 87, 142

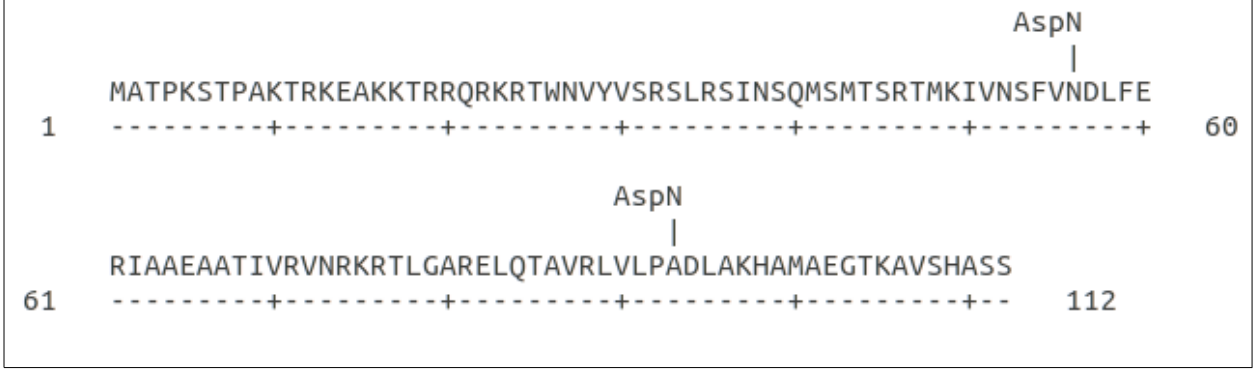
Supplementary document R1 (continued):

Histone H2B:

Name of enzyme	No. of cleavages	Positions of cleavage sites
Glu-C	5	13, 59, 64, 81, 101



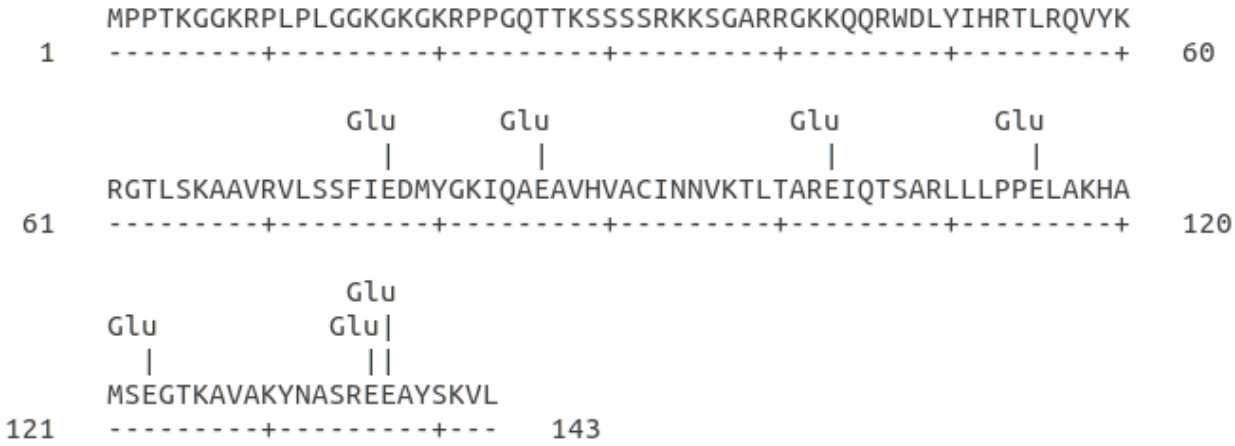
Name of enzyme	No. of cleavages	Positions of cleavage sites
Asp-N	2	56, 93



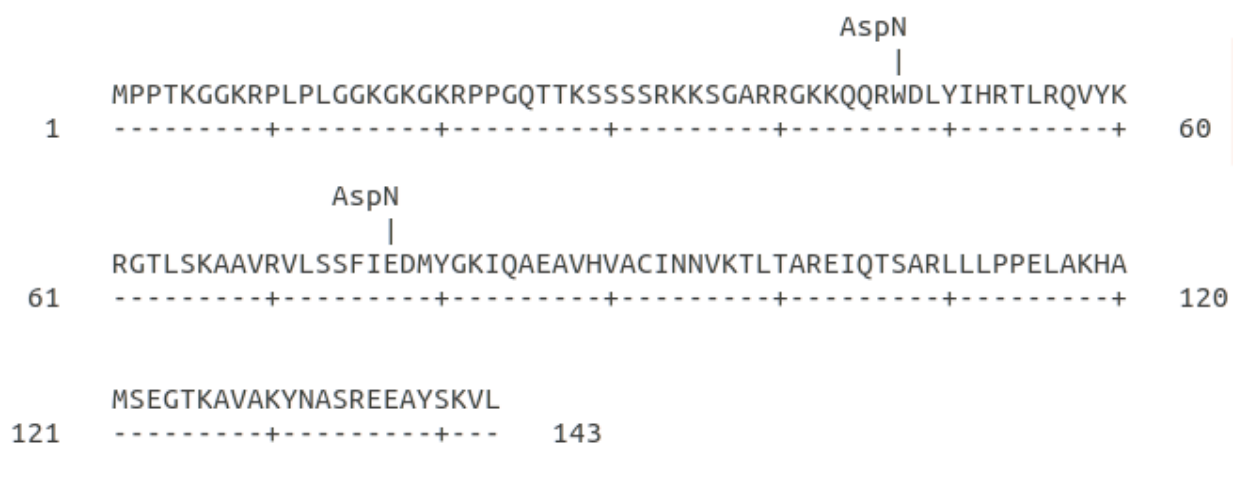
Supplementary document R1 (continued):

Histone H2B.V:

Name of enzyme	No. of cleavages	Positions of cleavage sites
Glu-C	7	77, 86, 103, 115, 123, 136, 137



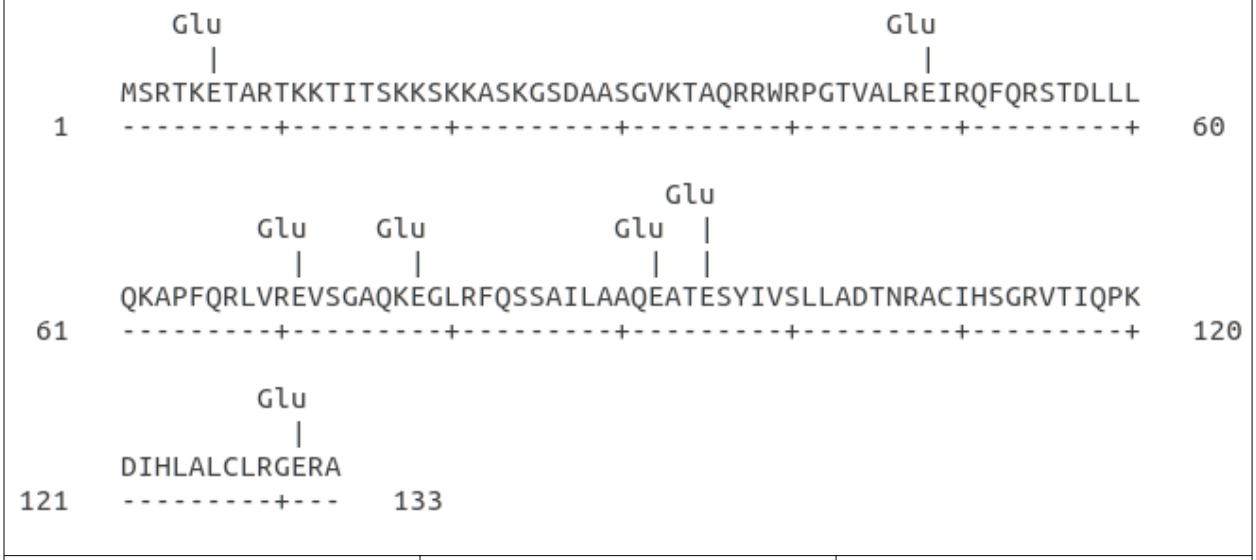
Name of enzyme	No. of cleavages	Positions of cleavage sites
Asp-N	2	44, 77



Supplementary document R1 (continued):

Histone H3:

Name of enzyme	No. of cleavages	Positions of cleavage sites
Glu-C	7	6, 48, 71, 78, 92, 95, 131

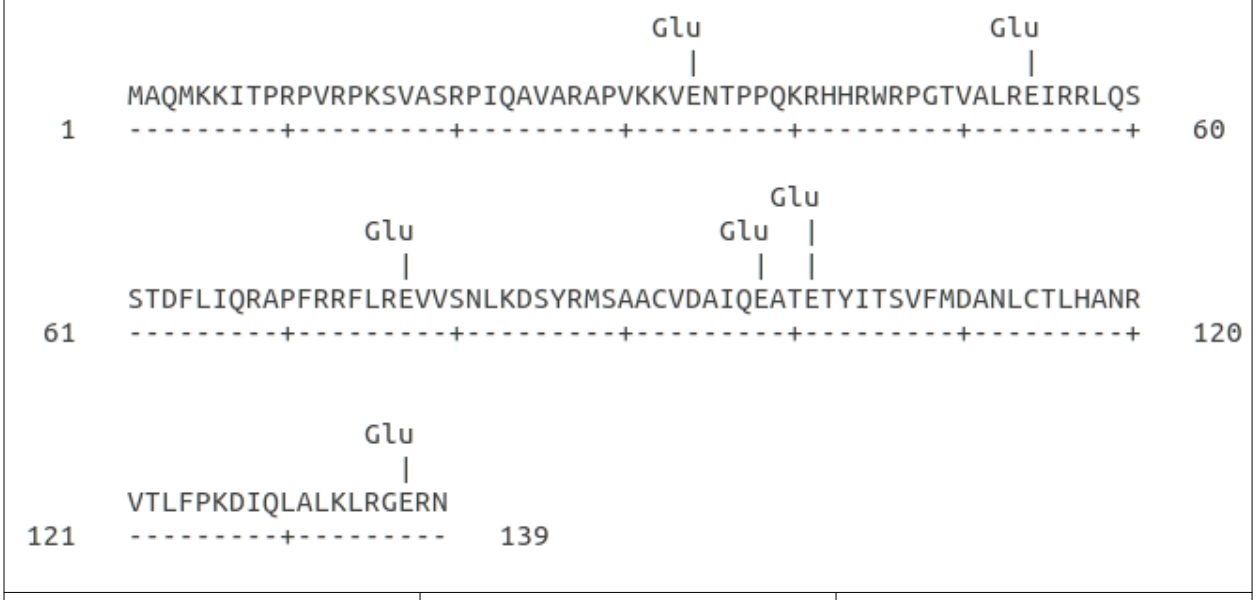


Name of enzyme	No. of cleavages	Positions of cleavage sites
Asp-N	4	26, 56, 103 ,120

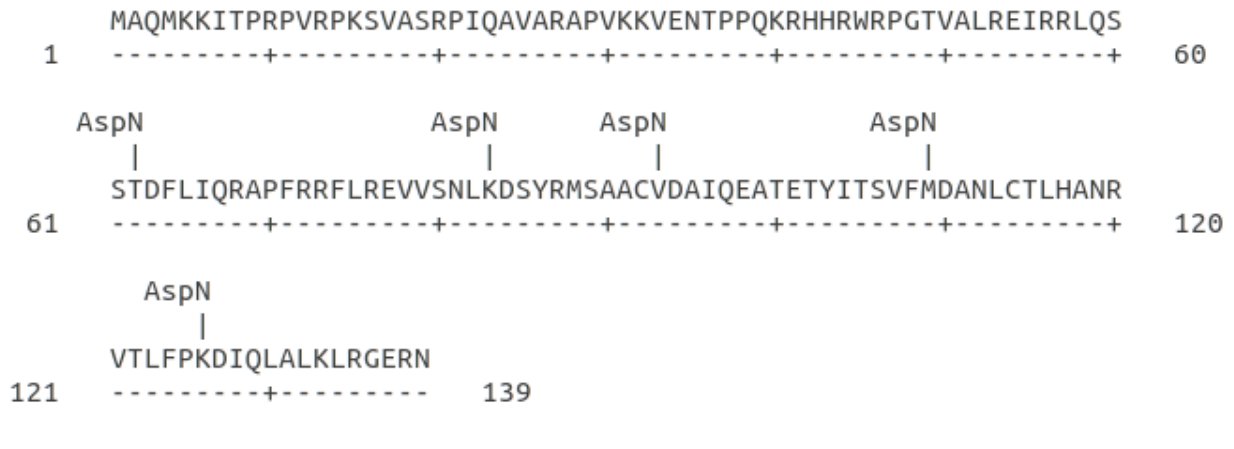
Supplementary document R1 (continued):

Histone H3.V:

Name of enzyme	No. of cleavages	Positions of cleavage sites
Glu-C	6	34, 54, 77, 98, 101, 137



Name of enzyme	No. of cleavages	Positions of cleavage sites
Asp-N	5	62, 83, 93, 109, 126



Supplementary document R1 (continued):

Histone H4:

Name of enzyme	No. of cleavages	Positions of cleavage sites
Glu-C	5	9, 23, 52, 62, 73

```
           Glu           Glu                           Glu
            |             |                             |
    MAKGKKSGEAKGSQKRQKKVLRENVRGITRGSIRRLARRGGVKRISGVIYDEVRGVLKSF
 1  ---------+---------+---------+---------+---------+---------+  60


   Glu        Glu
    |          |
    VEGVVRDATAYTEYSRKKTVTAVDVVNALRKRGKILYGYA
61  ---------+---------+---------+---------+  100
```

Name of enzyme	No. of cleavages	Positions of cleavage sites
Asp-N	3	50, 66, 83

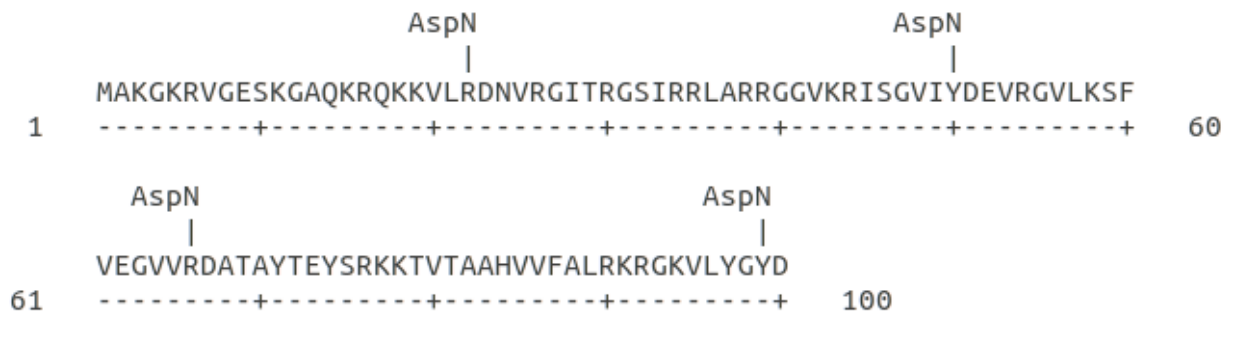
```
                                                    AspN
                                                     |
    MAKGKKSGEAKGSQKRQKKVLRENVRGITRGSIRRLARRGGVKRISGVIYDEVRGVLKSF
 1  ---------+---------+---------+---------+---------+---------+  60


      AspN             AspN
       |                |
    VEGVVRDATAYTEYSRKKTVTAVDVVNALRKRGKILYGYA
61  ---------+---------+---------+---------+  100
```

Supplementary document R1 (continued):

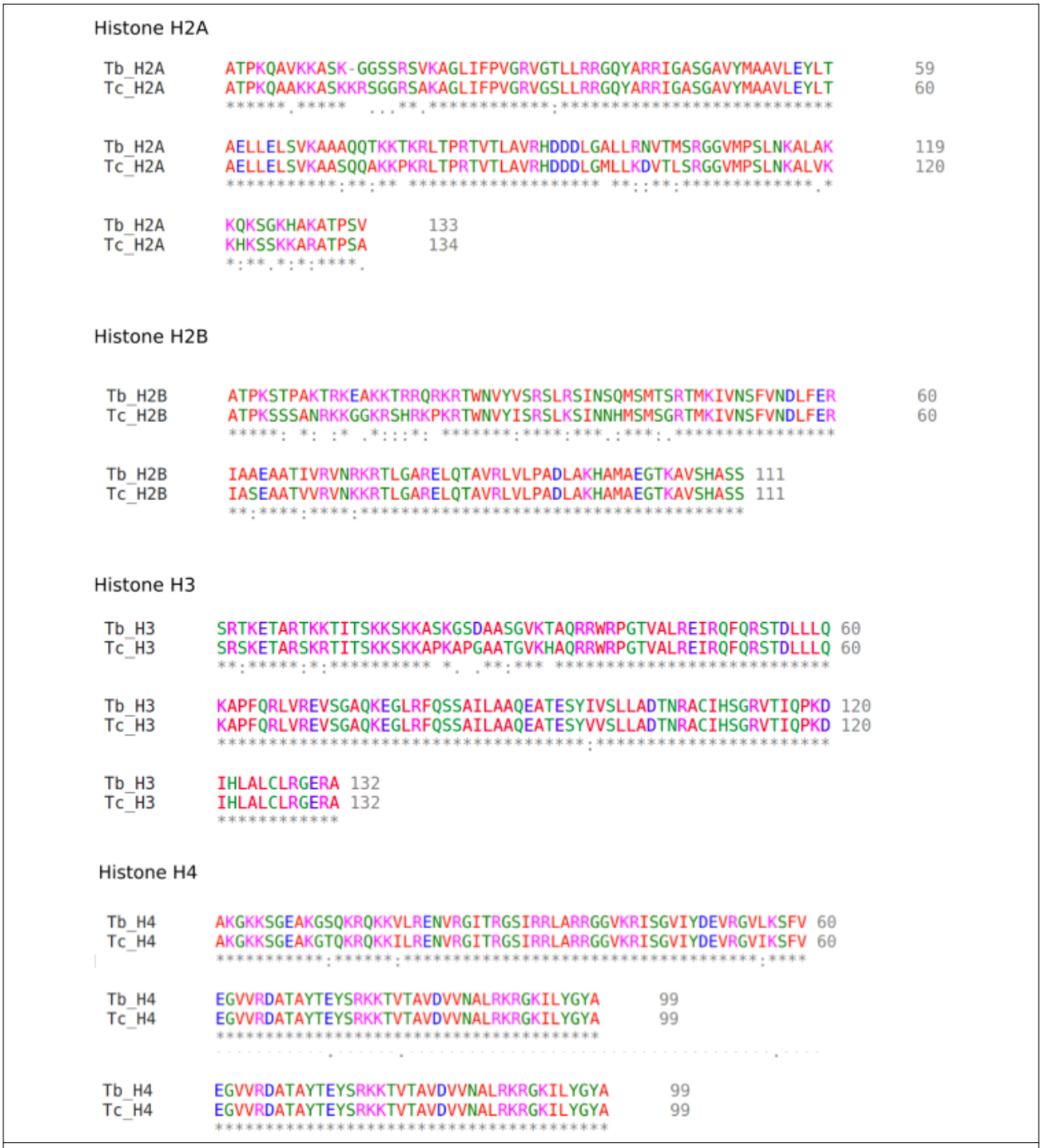
Histone H4.V:

Name of enzyme	No. of cleavages	Positions of cleavage sites
Glu-C	4	9, 52, 62, 73

Name of enzyme	No. of cleavages	Positions of cleavage sites
Asp-N	4	22, 50, 66, 99



Supplementary document R2:



Supplementary document R2: Sequence alignment of canonical core histones from *T. brucei* (Lister 427 strain) and *T. cruzi* (CL Brener strain).

Supplementary document R3: Protein sequences (with accession numbers underlined) retrieved from TriTrypDB (release 25, December 2015, www.http://tritrypdb.org) used to create a custom *T. brucei* Lister 427 Mascot searchable database.

T. brucei CANONICAL histones

>Tb_H2A | Tb427.07.2830

MATPKQAVKKASKGGSSRSVKAGLIFPVGRVGTLLRRGQYARRIGASGAVYMAAVLEYLTAELLELSVKAAAQQTKKTKRL

TPRTVTLAVRHDDDLGALLRNVTMSRGGVMPSLNKALAKKQKSGKHAKATPSV

>Tb_H2B | Tb427.10.10460

MATPKSTPAKTRKEAKKTRRQRKRTWNVYVSRSLRSINSQMSMTSRTMKIVNSFVNDLFERIAAEAATIVRVNRKRTLGAR

ELQTAVRLVLPADLAKHAMAEGTKAVSHASS

>Tb_H3 | Tb427.01.2430

MSRTKETARTKKTITSKKSKKASKGSDAASGVKTAQRRWRPGTVALREIRQFQRSTDLLLQKAPFQRLVREVSGAQKEGLR

FQSSAILAAQEATESYIVSLLADTNRACIHSGRVTIQPKDIHLALCLRGERA

>Tb_H4 | Tb427.05.4170

MAKGKKSGEAKGSQKRQKKVLRENVRGITRGSIRRLARRGGVKRISGVIYDEVRGVLKSFVEGVVRDATAYTEYSRKKTVTA

VDVVNALRKRGKILYGYA

Supplementary document R3 (continued):

T. brucei VARIANT histones

>Tb_H2A.V | Tb427.07.6360

MSLTGDDAVPQAPLVGGVAMSPEQASALTGGKLGGKAVGPAHGKGKGKGKGKRGGKTGGKAGRRDKMTRAARADLNF
PVGRIHSRLKDGLNRKQRCGASAAIYCAALLEYLTSEVIELAGAAAKAQKTERIKPRHLLLAIRGDEELNQIVNATIARGGVVP
FVHKSLEKKIIKKSKRGS

>Tb_H2B.V | Tb427tmp.02.5250

MPPTKGGKRPLPLGGKGKGKRPPGQTTKSSSSRKKSGARRGKKQQRWDLYIHRTLRQVYKRGTLSKAAVRVLSSFIEDMYG
KIQAEAVHVACINNVKTLTAREIQTSARLLLPPELAKHAMSEGTKAVAKYNASREEAYSKVL

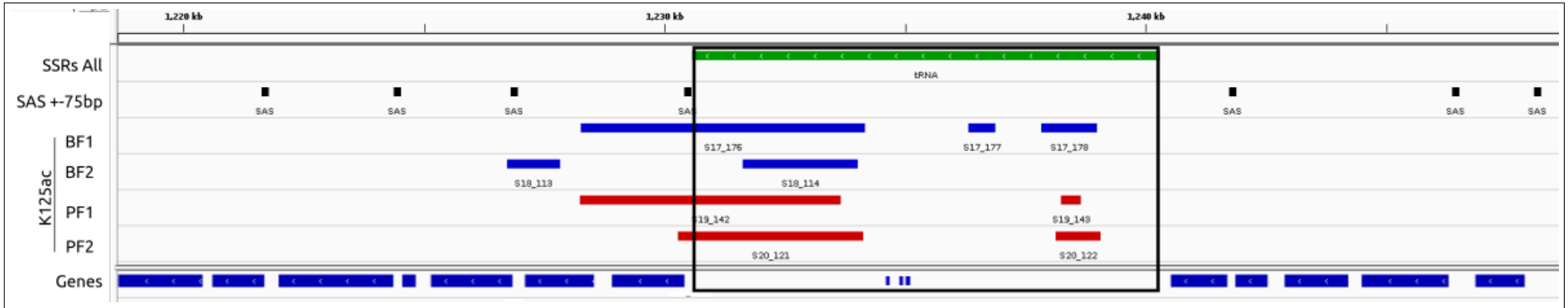
>Tb_H3.V | Tb427.10.15350

MAQMKKITPRPVRPKSVASRPIQAVARAPVKKVENTPPQKRHHRWRPGTVALREIRRLQSSTDFLIQRAPFRRFLREVVSN
LKDSYRMSAACVDAIQEATETYITSVFMDANLCTLHANRVTLFPKDIQLALKLRGERN

>Tb_H4.V | Tb427.02.2670

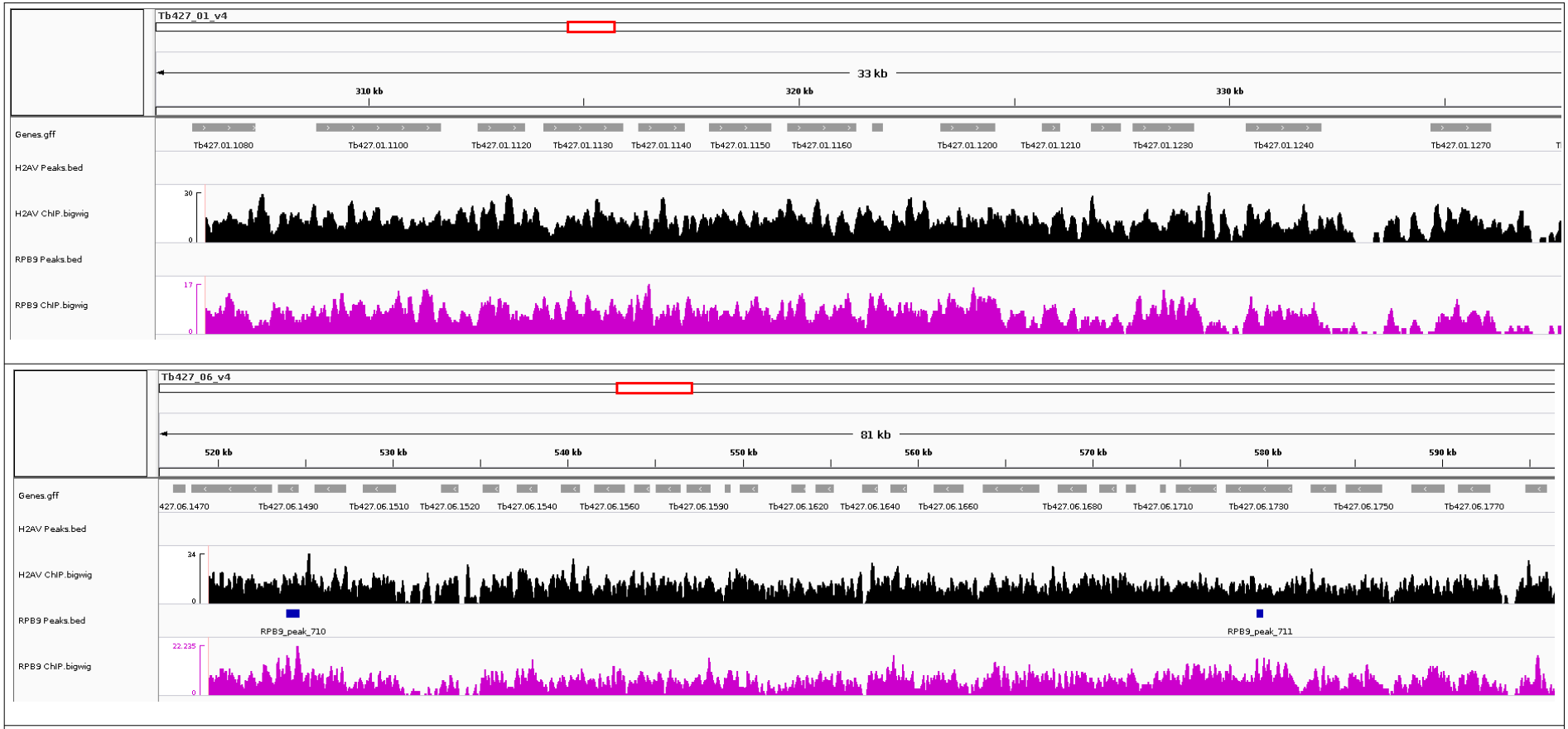
MAKGKRVGESKGAQKRQKKVLRDNVRGITRGSIRRLARRGGVKRISGVIYDEVRGVLKSFVEGVVRDATAYTEYSRKKTVTA
AHVVFALRKRGKVLYGYD

Supplementary document R4:



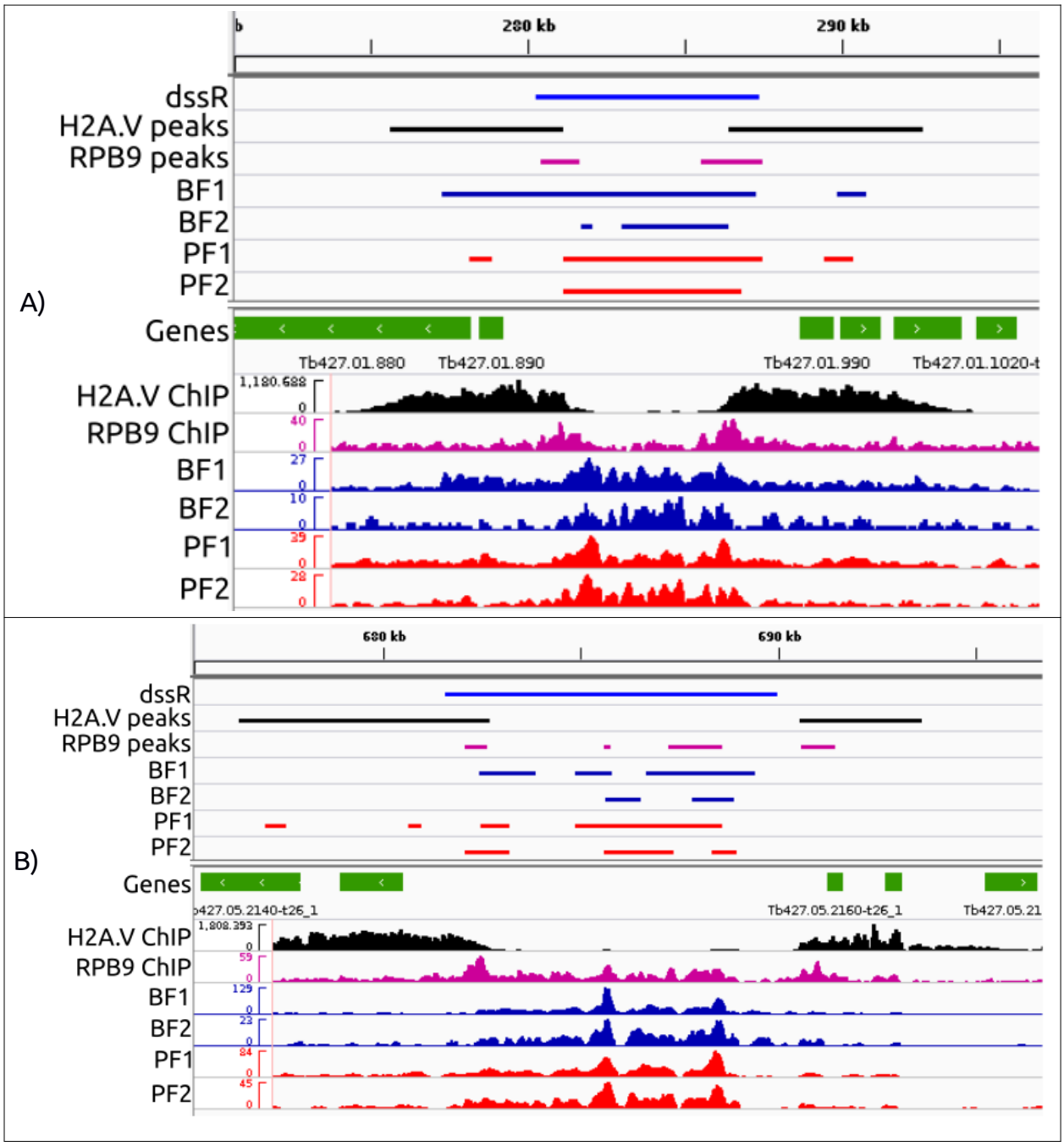
Supplementary document R4: K125ac peaks present at an internal stop/start caused by PTU interruption by a tRNA gene. ISSs and SASs are indicated by green and black bars in the first 2 tracks. Next, all K125ac peaks are shown (BF dark blue, PF in red). ISS region is indicated by the black box.

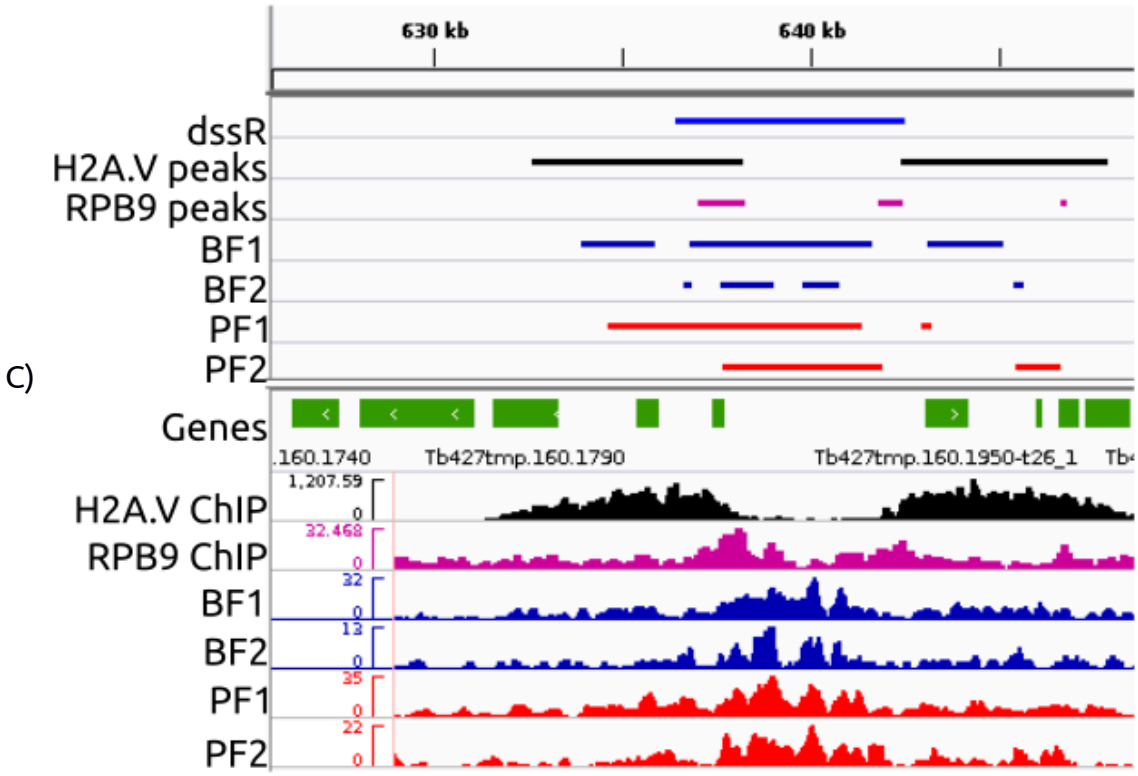
Supplementary document R5:



Supplementary document R5. H2A.V and pol II detected at basal levels across the entire PTU. Histone H2A.V (black) and RPB9 (magenta) ChIP profiles extend into the PTU and can be detected at low levels across entire pol II PTUs. (A) Chromosome 1 (position 305075 - 336925) and (B) chromosome 6 (position 516597 - 598952) is shown as representatives.

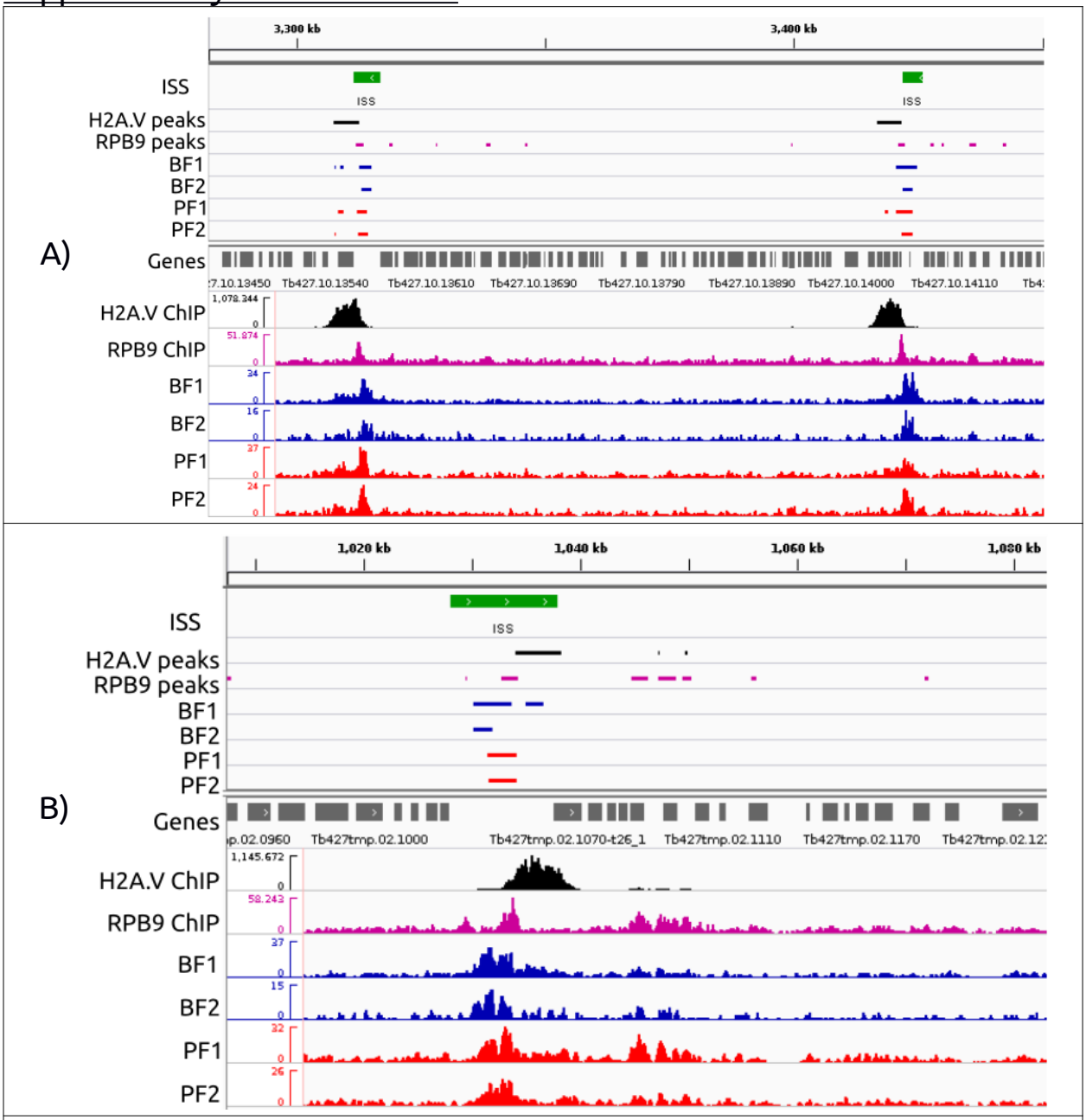
Supplementary document R6A:





Supplementary document R6A. H2A.V, RPB9, and K125ac peaks (BF1, 2 and PF1, 2) and ChIP profiles at dSSRs on (**A**) chrom 1 (position 262534 - 298882), (**B**) chrom 5 (position 670697 - 696458), and (**C**) chrom 9 (position 617981 - 652984) showing peak enrichment of H2A.V (black), RPB9 (magenta) and K125ac samples (BF samples 1, 2 in dark blue; PF samples 1, 2 in red) at divergent strand switching regions (top track, light blue). Genes are shown in green. H2A.V, RPB9, and K125ac ChIP profiles using the same color schemes as K125ac peaks

Supplementary document R6B:



Supplementary document R6B. H2A.V, RPB9, and K125ac peaks (BF1, 2 and PF1, 2)and ChIP profiles at internal stop/start sites on (**A**) chromosome 10 (position 3282255 - 3545768) and (**B**) chromosome 11 (position 989909 - 1066791)> H2A.V (black) peaks can be seen immediately downstream of ISSs (green). RPB9 (magenta), K125ac samples BF (1, 2; dark blue), and PF (1, 2; red) replicates localizing on the ISSs. Genes are indicated in gray. PTU directionality is indicated by the arrows in the green ISS blocks.